

AUTOMATIC HIGH-SPEED COMPUTING

A Progress Report on the EDVAC

Report of Work under Contract No. W-670-ORD-4926,
Supplement No. 4

Between

Ordinance Department, United States Army
Washington, D. C.

and.

The University of Pennsylvania,
Moore School of Electrical Engineering
Philadelphia, Pa.

This document contains information affecting
the national defense of the United States
within the meaning of the Espionage Act, U.S.
C. 50: 31 and 32. Its transmission or the
revelation of its contents in any manner to
an unauthorized person is prohibited by law.

This Report Was Prepared by:

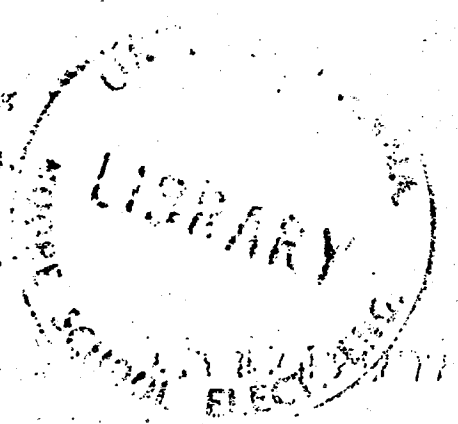
J. Prosper Eckert, Jr.
John W. Mauchly

Moore School of Elect. Eng.

This is copy No. 9
of 50 bound copies of
this report.

MOORE 510.78/P387.9/cop. 2

M D Godfrey
5/83
Unclassified
per GLE Dear Bully
OCOD
CAGT Powell
"Mar 47



ACKNOWLEDGEMENTS

The authors of this report are greatly indebted to Dr. Harry Huskey, who wrote most of parts 5, 6, and 7 and worked out the block diagrams. His contributions to the planning of this report are also appreciated.

Acknowledgement is also due those who, from time to time, have participated in discussions of EDVAC plans. Those who have contributed in this way are Dr. Arthur Burks, of the Moore School Staff, Capt. H. H. Goldstine, Army Ordnance, and especially Dr. John von Neumann, consultant to the Ballistic Research Laboratory.

September 30, 1945

J. P. Eckert, Jr.
J. W. Mauchly

The engineering staff for the development of the EDVAC includes, in addition to those mentioned in the "Acknowledgements" above,

Joseph Chedaker
Chuan Chu
James A. Cummings
John H. Davis

F. Robert Michael
T. Kite Sharpless
Robert F. Shaw
C. Bradford Shppard

S. Reid Warren, Jr., Project Supervisor

Pennsylvania University: Moore school of electrical engineering

Moore School

TABLE OF CONTENTS

Part		Page
1	INTRODUCTION AND SUMMARY	1
1.0	Historical Comments	1
1.1	Contrast of the EDVAC with the ENIAC	4
1.2	Summary of Report	6
2	GENERAL DISCUSSION OF COMPLEX COMPUTING MACHINES	10
2.0	Purpose of this Part	10
2.1	Computing Machines as Tools	10
2.2	Model Experiments	12
2.3	Analogy Devices and Network Analyzers	13
2.4	The Differential Analyzer	15
2.5	The Argument for Serial Operation	17
2.6	Electronic Devices Needed for Speed	18
2.7	Program Information and Other Inputs	19
2.7.0	Classification of Input Information	19
2.7.1	Inputs to Analogy Machines	20
2.7.2	Inputs to Arithmetic Calculators	21
2.7.3	Speed of Input and Control Devices	23
2.8	Accuracy Requires Digital Computing	25
2.9	The Economics of Computing Machines	27
2.9.0	The Cost of Mathematical Solutions	27
2.9.1	Relative Costs for Memory	30
2.9.2	Comparison of Input and Output Devices	35
2.9.3	Graphical Versus Tabular Outputs	36
2.9.4	Preparation of the Problem	38
3	THE DELAY LINE MEMORY DEVICE	41
3.0	The Delay Line Memory Device	41
3.1	Operation of the Delay Line Memory	41
3.2	Equipment and Storage Capacity	45
3.3	Terminology for EDVAC Memory	46
3.4	Short versus Long Tanks	47
4	MAGNETIC INPUT AND OUTPUT DEVICES	48
4.0	General Features	48
4.1	Recording from High-speed Memory	49
4.2	Reading into High-speed Memory	51
4.3	Reversibility of Wire Direction	52
4.4	Compactness of Wire or Tape Records	53
4.5	Sorting and Collating	54
4.5.0	The Need for Sorting and Collating	54
4.5.1	Tape versus Card Sorting	55
4.5.2	A Binary Sorting Device	57
4.5.3	Sorting Speed	59
4.5.4	Systematic Sorting	61
4.5.5	Collating	61
4.6	Servo Control for Magnetic Reading	63
4.7	Typewriters	63

TABLE OF CONTENTS (cont.)

Part		Page
5	BRIEF DESCRIPTION OF EDVAC COMPONENTS	66
5.0	General Structure of the EDVAC	66
5.1	Delay Line Memory Devices (Tanks)	66
5.2	Electronic Switches	67
5.3	Computer	68
5.4	Controller	68
5.5	Magnetic Input and Output Devices	68
5.6	Oscilloscope for Graphical Output	69
5.7	Power Supplies	70
5.8	Auxiliary Devices	70
6	NUMBER AND ORDER CODES	71
6.0	General Considerations	71
6.1	Lengths of Words and Numbers	71
6.2	Coded Decimal Systems	73
6.3	The Order Code	75
7	TENTATIVE PLANS FOR THE EDVAC	77
7.0	Number and Arrangement of Components	77
7.1	Estimate of Tube Requirements	82
7.2	Suggested Switch Connections	83
7.3	Pulse Code for Orders	83
7.4	Control	86
7.5	Delay Line Memory Circuit	88
7.6	The Computer	91
7.6.0	The Computer	91
7.6.1	The Binary Adder	91
7.6.2	The Coded Decimal Adder	93
7.6.3	Negative Numbers	93
7.6.4	Multiplication	95
7.6.5	Division	96
7.6.6	Comparing	96
7.6.7	Significant Figures and the Decimal Point	97
8	RELIABILITY AND CHECKING	98
8.0	Character of Failures	98
8.1	Types and Levels of Checking	102
8.2	Continuous Versus Intermittent Checking	103
8.3	Checking in Serial Machines	104
8.4	Checking in Other Machines	105
8.5	Checking the Computer and Control Units	106
8.6	Checking the Memory and Switches	108
8.7	Checking the Input and Output	110
8.8	Smoothness Checking and Programmed Checking	110
9	CONCLUSION	111

CONFIDENTIAL

CONFIDENTIAL

LIST OF FIGURES

Fig. No.		Page
1	Simplified Block Diagram for Delay Line Memory Device	42
2	Approximate Pulse Shape	42
3	EDVAC Plan A	79
4	EDVAC Plan B	80
5	Memory Tank Circuit	89
6	Decimal Adder	94
7	Electronic Switch	90
PY-2-101	Binary Pulse Adder	92

CONFIDENTIAL

AUTOMATIC HIGH SPEED COMPUTING

A Progress Report on the EDVAC

Part 1 INTRODUCTION AND SUMMARY

1.0 HISTORICAL COMMENTS

The Ballistic Research Laboratory of the Aberdeen Proving Ground is forced, by the nature of its work, to carry out numerical computations which are notable both for their volume and for the variety and complexity of the procedures employed. Of necessity, the Laboratory has become an important computing center. To expedite its work, it has sought improved techniques for the use of existing equipment and has fostered the development of new devices for "mechanizing" and speeding up the processes of numerical computation.

In August, 1942, J. W. Mauchly summarized briefly, in memorandum form, the advantages to be expected from an electronic high-speed computer such as could reasonably be developed at the time. Capt. Herman H. Goldstine, of the Ballistic Research Laboratory, and Col. Paul N. Gillon, of the Office of the Chief of Ordnance, became interested in the possibilities of such a device for the Laboratory. Mauchly and J. P. Eckert, Jr. then wrote out a tentative technical outline of a machine which would be capable of numerically integrating trajectories for firing tables and which would handle other computing jobs of similar complexity. J. G. Brainerd then incorporated this material into a report which was submitted to the Ballistic Research Laboratory. The United States Government then entered into a contract with the University of Pennsylvania, to have the Moore School of Electrical Engineering carry out research and development on electronic calculating devices for the Ballistic Research Laboratory. The project was set up under the administrative supervision of Dr. Brainerd with Mr. Eckert as chief engineer and Dr. Mauchly as principle consultant; Captain Goldstine was appointed to take technical cognizance of the project for the Laboratory.

CONFIDENTIAL

Under this contract, a staff at the Moore School have, since July 1, 1943, worked continuously on the design and construction of electronic computing devices. A first machine, the Electronic Numerical Integrator and Computer (ENIAC) has been designed and constructed, and will soon be put into operation. It was necessary, in order to complete a task of this magnitude in little more than two years, to fix the design at an early date and to concentrate thereafter on production and testing.

During this latter period of work on the ENIAC, it became more and more evident that close scrutiny should be given to the way in which the high speed of electronic devices could be utilized in computing machines. It became apparent that serial operation was in general advantageous, and that when serial methods were used wherever possible the equipment was used most efficiently. Hence, in January, 1944, a "magnetic calculating machine" was disclosed, wherein the successive digits of a number were transmitted in timed sequence from magnetic storage or memory devices through electronic switches to a central electronic computing circuit and similarly returned to magnetic storage. An important feature of this device was that operating instructions and function tables would be stored in exactly the same sort of memory device as that used for numbers. The electronic switching means for such a machine were already disclosed -- the ENIAC function tables use one form of such a switch.

The invention of the acoustic delay line memory device by Eckert and Mauchly early in 1944 provided a way of obtaining large high-speed storage capacity with comparatively little equipment. An automatic electronic calculating machine was then planned, using the delay line device for the "internal" memory, and using the somewhat lower speed magnetic devices only for input and output mechanisms.

It was clear that this new machine would, with much less equipment, easily handle problems beyond the intended scope of the ENIAC. Therefore, by July, 1944 it was agreed that when work on the ENIAC permitted, the development and construction of such a machine should be undertaken. This machine has come to be known as the EDVAC (Electronic Discrete Variable Computer).

Although pressure of work on the ENIAC has made it impossible to conduct much experimental work on the EDVAC, all of the essential components to be used in the EDVAC are known to work. In designing a specific machine, however, a great number of decisions must be made -- decisions as to quantity and proportion of various parts, and decisions as to which of several possible methods shall be employed. During the past year, many questions of this sort have been considered in a series of discussions held just for this purpose.

During the latter part of 1944, and continuing to the present time, Dr. John von Neumann, consultant to the Ballistic Research Laboratory, has fortunately been available for consultation. He has contributed to many discussions on the logical controls of the EDVAC, has proposed certain instruction codes, and has tested these proposed systems by writing out the coded instructions for specific problems. Dr. von Neumann has also written a preliminary report* in which most of the results of earlier discussions are summarized. In his report, the physical structures and devices proposed by Eckert and Mauchly are replaced by idealized elements to avoid raising engineering problems which might distract attention from the logical considerations under discussion.

*First Draft of a Report on the EDVAC: June 30, 1945

It must be emphasized that final decisions on many questions cannot yet be made. In many cases, closer study of alternative circuit designs of further experimental data will be necessary before intelligent choices can be made. The discussions that have been held have served an essential purpose by formulating the important questions and indicating what experimental data is pertinent to their answers.

1.1 CONTRAST OF THE EDVAC WITH THE ENIAC

The EDVAC and the ENIAC have very little in common. Although they are both electronic digital computing machines, the resemblance ends there. A few salient contrasts will emphasize this. In many respects the ENIAC is an electrical analogue of mechanical computing machines: it has decade counters instead of decade gear wheels, and all the decimal digits of a given number are added simultaneously as in any mechanical adding machine. Multiple channels and multiple equipment are necessary to do this. In the EDVAC, the various digits specifying a number will be transmitted serially through a single channel. Switching equipment is consequently simplified and reduced. Moreover, the storage or memory elements to be used in the EDVAC are entirely different from those used in the ENIAC. A typical unit, involving about 10 vacuum tubes, can be made to hold many digit sequences in definite time relationship.

In fact, the EDVAC will have an internal high-speed memory capacity for about 2,000 ten-digit decimal numbers, while the ENIAC accommodates only 20. Nevertheless, the EDVAC is expected to require only one-tenth as much equipment as the ENIAC. As a consequence, the solution of wide classes of non-linear partial differential equations will be easily managed in the EDVAC, whereas such problems can be carried out on the ENIAC only with the aid of a much slower external memory capacity in the form of punched cards.

Another important contrast between the ENIAC and the EDVAC is this: To change over from one type of problem to another, the ENIAC must be prepared by setting numerous switches and plugging in many interconnecting cables, while the EDVAC will be set up automatically and electronically for each new problem. Indeed, the EDVAC will store its operating instructions in exactly the same manner as it stores numbers. The total storage capacity will be at the disposal of the operator to allocate as he wishes--part to operating instructions, part to tables of functions, and part to the numbers which enter into and come out of the various calculations.

The input and output devices for the two machines are also different. The ENIAC reads from IBM cards, and results are punched into IBM cards. It is expected that the EDVAC will read from a steel wire or tape on which the required information has been recorded by magnetic means, and that the output data will be magnetically recorded on steel wire. Typewriters for preparing such wire and for printing from wire records will be provided. Records on steel wire are less bulky and ^{more easily and} automatically handled than stacks of separate cards. Also, higher reading and recording speeds are attainable with steel wire than with cards or with paper tape. Sorting problems, for which cards have always been considered desirable, are also done expeditiously using wire.

The EDVAC will be, then, a mathematical tool of much greater power and scope than the ENIAC. At the same time it will be physically smaller and easier to service.

Finally, much attention has been given to the question of "coding the problem" so as to simplify the work of the mathematician who must "instruct" the machine as to what he wants done in symbols which the machine "understands". On this topic and many others the project staff at the Moore

School has benefited greatly from conferences with and work done by Dr. John von Neumann, acting as consultant to the Ballistic Research Laboratory.

In Table 1 are some comparative figures for the ENIAC and the EDVAC. The figures for the EDVAC are estimated on the basis of the tentative plans outlined in Part 7 of this report.

1.2 SUMMARY OF REPORT

The earlier sections of this report are given over to a general discussion of complex computing machines, with emphasis on the way in which such machines employ memory devices, serial scanning procedures, and programming methods. This approach indicates that a machine of the necessary generality, capacity, and accuracy can be realized without excessive and unreasonable amounts of equipment if (a) "scanning" is employed so that one piece of equipment can be used many times rather than having many similar components working simultaneously, (b) the speed lost by this serial process is regained by employing electronic elements, and (c) accuracy is insured by digital methods, as, for instance, by the use of vacuum tubes as on-off devices, rather than as continuous variable devices.

Although serial use allows computing and switching equipment to be reduced to a minimum, the large storage capacity which is essential for numerical data, operating instructions and function tables would involve prohibitive amounts of electronic equipment if decade rings, binary counters, or other previously known methods were employed. A new and inexpensive memory device, the delay line register, conceived by J. P. Eckert and J. W. Mauchly in 1943, makes it possible to realize with a small amount of equipment the memory capacity needed for the EDVAC. The essential features and operation of this device are explained in Part 3.

TABLE I

Comparative figures for the ENIAC and the EDVAC

NOTE: As explained in the text, the work of arranging or coding problems for solution by the EDVAC will be simpler than for the ENIAC, and the physical set-up of the EDVAC will be automatic, while there are 5000 manual switches and hundreds of cords to be plugged in manually in order to set up the ENIAC.

	ENIAC	EDVAC
<u>MEMORY CAPACITY</u> (each unit is one nine or ten digit decimal number, or one program instruction)		
Auxiliary memory	Unlimited, on IBM cards	Unlimited, on magnetic wire or tape
Internal high speed memory	20	Over 2000
Manually set up programs and function tables	725	0
<u>SPEED OF OPERATION</u> in milliseconds (9 or 10 digit decimal numbers)		
Addition	0.2	0.06
Multiplication	2.6	1.0
Input from auxiliary memory	35	3 to 6
Output to auxiliary memory	70	3 to 6
<u>EQUIPMENT</u>		
Vacuum tubes	18,000	1,925
Manual switches	5,000	Possibly a few
Digit and program channels	200	2
Power supplies	28	3
<u>SPACE REQUIREMENTS</u> (Sq. ft.)		
Floor area actually occupied by machine	250	25
Minimum area of room to allow access to machine	1500	125

The high computing speed of the EDVAC would be largely wasted if conventional input and output devices were used. A new method, using magnetic wire or tape, has therefore been examined and found to meet EDVAC requirements. Input and output devices of this type are described in Part 4. A discussion of sorting and collating is included to show that magnetic wire devices are better for this purpose than existing card machines. Typewriters for preparing magnetic records and for printing from magnetic records are also discussed.

Part 5 contains a brief description of all of the principal components which will be used in the EDVAC. These are the memory units, the computer and the controller, electronic switches, the magnetic input and output devices, an oscilloscope unit for visual output, the power supplies and auxiliary devices such as typewriters.

Some desiderata for coding are presented in Part 6, which also contains an instruction code suggested by von Neumann and a proposed numerical code.

Plans for the EDVAC are presented in more detail in Part 7. These plans are tentative, and will be modified as the design is advanced. For the purpose of this report, it was decided to present plans based on use of a coded decimal number representation, although a binary system may ultimately be used. The computer elements are therefore somewhat more complicated than they would be for a binary machine. Two different plans for switching connections and corresponding pulse codes for orders are given to illustrate ways of organizing the components of the EDVAC.

In part 8 the reliability of electronic components and the methods which should be used for detecting failures in the EDVAC are given careful consideration. Until electronic digital computers are in use, and

CONFIDENTIAL

pertinent failure statistics become available, it is difficult to know how elaborate checking methods must be. It is therefore proposed that the EDVAC have a rather complete checking system which will guarantee the detection of almost all failures. This system checks all standard operations automatically, without instructions from the operator. Since the operator may always instruct the machine to perform further mathematical checks, full confidence in the accuracy of EDVAC solutions is assured.

In conclusion, it is intended, as stated in Part 9, that additional reports will be issued to supplement this one, and that some of these reports will give examples of how specific problems would be handled by various proposed instruction codes.

CONFIDENTIAL

PART 2. GENERAL DISCUSSION OF COMPLEX COMPUTING MACHINES

2.0 PURPOSE OF THIS PART

In the following sections under this part, a general discussion of various complex computing machines will be used to emphasize the way in which such machines employ memory devices, serial scanning procedures, and programming methods. This approach indicates that a machine of the necessary generality, capacity and accuracy can be realized without excessive equipment if (a) "scanning" is employed so that one piece of equipment can be used many times rather than having many components working simultaneously, (b) the speed lost by this serial process is regained by employing electronic elements, and (c) accuracy is insured by digital methods, as for instance, by the use of vacuum tubes as on-off devices rather than as continuous variable devices. The EDVAC is to have the form which this general discussion indicates as desirable.

2.1 COMPUTING MACHINES AS TOOLS

For many branches of science and engineering, mathematics is a basic tool. A great deal of progress has been made in developing this tool in a formal way. Archimedes and Galileo spoke in terms of simple proportionalities; today many relationships are expressed in terms of integral and partial differential equations.

For engineering design, however, numerical results are needed. Also, the scientist cannot test his theories unless he can make numerical predictions to compare with experiment. How are numerical results to be obtained from the basic mathematical theory?

There is no perplexity as to the meaning of the mathematical symbols, and in most cases, a numerical procedure for developing the required solution to a good approximation is known. But the computation by these direct step-by-step methods becomes complex and voluminous even for relatively simple ordinary differential equations. Although ordinary desk computing machines are tools of great value and usefulness, more complicated tools, such as the differential analyzer and the network analyzer, must be brought to bear upon these problems. Although these devices greatly extend the field in which numerical solutions may be achieved, there exist today many important engineering and scientific problems which fall outside the scope of these machines.

The equations appropriate to such problems are known, and can easily be written out. Sometimes a formal solution can be exhibited, but calculation of numerical results from this formal solution (for example, a definite integral) remains exceedingly difficult and tedious. In many cases, the theory involves a non-linear partial differential equation or an integral equation which does not fall into one of the well-known types, and even a formal solution in terms of known functions cannot be obtained. In this event, all sorts of remedies are tried. Special cases are sought, awkward terms are dropped out, and dubious approximations are resorted to, in an effort to render the equations tractable. A great deal of effort is spent in expanding solutions in terms of known analytic functions. Naturally, solutions obtained in this way may be exceedingly restricted, or of doubtful applicability. Even though a formal solution is obtained, there remains the problem of numerical evaluation which can be exceedingly difficult for slowly convergent series.

It may be questioned whether, in such circumstances, the game is worth the candle. Can't the desired answers be secured in some other way? For engineering purposes, a model experiment may sometimes furnish the necessary data. But no model experiment can serve to test the adequacy of a scientific theory. The sole question at issue is then whether numerical predictions from the theoretical relationships agree with experiment. And until the theory can be tested by this comparison, further advance in the theory is slowed if not halted. A calculating machine capable of handling problems beyond the scope of existing machines is therefore a scientific tool of first importance.

In engineering, whether the object to be designed is a single suspension bridge or a small projectile to be manufactured in large quantities, there is ample justification for intense design effort. For many cases, the general mathematical theory which governs the design is well known, but this theory is actually used only as a qualitative guide because of computing difficulties. A computing machine which could overcome these difficulties would be a most important tool. Even if the mathematical design were no cheaper, it would be desirable if it were quicker. Timeliness has both an economic and a strategic value.

2.2 MODEL EXPERIMENTS

From the point of view of this report, an engineer who makes measurements on a small scale model of the device which he intends to build, is using the model as a calculating device. This calculating device is exceedingly specialized. If he wishes to investigate the effect of changing some design parameter, more often than not he will have to have a series of models built, and a great many measurements will have to be made.

The model may be expensive, the measurements may be difficult, or require an expensive auxiliary such as a wind tunnel, and he may also have additional problems concerned with the effects which the measuring instruments themselves introduce into the experiment. But all these things he will put up with if no other method is available for providing the quantitative data he needs more quickly or more cheaply.

2.3 ANALOGY DEVICES AND NETWORK ANALYZERS

Quantitative solutions to specific problems may be obtained by resort to all sorts of analogy devices. For instance, a problem concerning electric fields may be solved by an electrolytic tank in which a probe is used to follow the streamlines of the electric current. Often electrical network analogies are used for the solution of problems arising in mechanical systems.

Sometimes electrical networks having discrete meshes are employed for the solution by analogy of problems concerned with continuous media. A resistance-capacity network is useful for the study of the flow of heat through a continuous medium. Even though the number of meshes in the network is not very large and, as a consequence, the data obtained from the network approximate the desired results only roughly, nevertheless the time required to obtain such solutions by this method may be small compared to the time required for other available methods of solution. Sometimes the networks are made from fixed electrical components and are thus good for only one specific problem. In order to facilitate their application to the solution of a wider range of problems, such networks may be constructed from adjustable components. Further means for establishing various interconnections between different elements may be provided. A device such as

this operated with alternating current is sometimes known as an a-c calculating board.

It is customary in using such a network to apply voltages or currents to the terminals of the network — which has previously been set up to represent by analogy a particular problem — and then to read other voltages or currents in the network which correspond to the desired results. Readings may be taken at a great number of points in the network, but usually only one or two instruments are used in doing this. For instance, the same voltmeter may be moved from point to point. It is, then, proper to say that the various points in the network are scanned to obtain a succession of numerical results.

It will be noted that the network analyzer is something like a complex model, the parameters of which can easily varied and on which many measurements can readily be made. However, the elements of such a network are usually linear elements, and the network is therefore best adapted to the solution of problems which can be expressed in terms of linear equations. If non-linear equations are to be solved, it is usually necessary to introduce non-linear elements into the network. Although many non-linear circuit elements are known, it is not easy — and in specific cases it may prove impossible — to find elements which approximate accurately enough the desired non-linear characteristics to satisfy the problem. Unfortunately, then, these problems for which a computing device is most needed are also the ones which are the most difficult to set up on a network analyzer.

2.4 THE DIFFERENTIAL ANALYZER

The differential analyzer is an analogy device capable of handling a large variety of non-linear differential equations. Each integrator is a non-linear element, and by suitable interconnection of two or more integrators, many non-linear relationships can be secured. Differential analyzers are also provided with "function tables", such as curve followers, capable of feeding in arbitrary functions, so that all non-linear effects need not be generated by the integrators. The differential analyzer can therefore be used to solve problems of a sort which cannot easily be solved on the network analyzer. The converse is also true, since to date it has not been practical to build differential analyzers having more than 10 or 20 integrators, and several function tables, while a network analyzer can have a great many more elements. The differential analyzer is not suitable, therefore, for the solution of a large number of simultaneous linear equations, while the network analyzer can be used for this purpose.

Several other contrasts between these two devices should be mentioned. As ordinarily used, the differential analyzer is a continuous variable machine, whereas the network analyzer provides data only at discrete points. Furthermore, the network analyzer (for a "steady state" set-up) provides all of the answers at once, even though these are customarily read off in sequence by scanning the network as explained above. The differential analyzer, on the other hand, has a "free parameter" which governs the development of the solution. This independent variable, often called the "time variable", controls the progress of the solution. As this variable is increased, selected dependent variables are developed by the machine and are plotted or printed. The desired results are thus taken out of the machine serially by a continuous scanning process.

In the network analyzer, only the meter is moved from point to point and the order of scanning is immaterial. In the differential analyzer, the independent variable "programs" the scanning process in a serial fashion. It is usual to think of the problem data being "fed through" the integrators and other elements of the analyzer, but we might equally well say that the limited number of non-linear devices at our disposal are being moved through the problem space. In any case, it appears that in order to utilize a limited number of special devices at a great number of points (in this instance, a continuum of points) a serial scanning process having one degree of freedom has been used.

In the network analyzer, the problem data are all fed in or set up prior to the scanning process. Information concerning the problem is "stored", so to speak, in memories consisting of switches and so forth which are not altered during the scanning process. In the differential analyzer, information pertinent to any stage of the calculation is stored in the form of angular shaft positions. The shafts are in fact memory elements, and the information stored in these memory elements is continuously altered as the problem is scanned.

The differential analyzer may be used, of course, to solve a set of simultaneous ordinary differential equations, all having the same independent variable, in such a manner as to approximate the solution of a partial differential equation. In this case, the solution is obtained in the form of a number of "simultaneous" discrete curves. In one variable, the results are continuous, as before, but in the other, they are discrete, as with the usual network analyzer. Since separate integrators must be used for each continuous integration path, it is not possible to integrate along very many paths simultaneously because of the small number of inte-

ters available. Because of this restriction in capacity, differential analyzers are not very effective in attacking partial differential equations.

2.5 THE ARGUMENT FOR SERIAL OPERATION

It is evident that problems having the complexity associated with non-linear partial differential equations cannot be solved by either the network analyzer or the differential analyzer in their present forms. Theoretically, a useful device would seem to need an enormous number of components, each capable of being given exceedingly arbitrary but well defined non-linear characteristics. There is, however, a very much more reasonable alternative. Suppose that only one such highly flexible and versatile non-linear element be provided, but that this element be used successively at each of the various "points" in a network analyzer. It would then be necessary, of course, to have a memory of some sort at each network point, so that once the non-linear element had established the proper relationship of variables at that point, this relationship would be maintained after the non-linear element were transferred to another point. By this serial use of a single non-linear device over and over again, the same results are ultimately achieved without the need for a large number of such non-linear elements. The construction of one (or possibly two) quite complex elements having the required flexibility, generality, and accuracy will be shown to be practical, whereas the cost, bulk and maintenance of large numbers of such elements would be prohibitive.

The network to be scanned by such a non-linear element is now a network of memory elements. There is no need for the electrical network in the original form of network analyzer. The large memory capacity needed for such a machine might seem to be impractical, but fortunately this is not so.

CONFIDENTIAL

In a later section it will be shown that such a memory capacity can easily be obtained with relatively simple equipment. Indeed, the simplicity of the memory system referred to will be found to be intimately associated with the fact that it, too, is serial in nature.

The principle of serial operation can be applied not only to the programming of the machine as a whole, but also to the operations which must occur within the various parts of the machine. For instance, the part which has been called a "general non-linear element" is in fact a computing device. Within it, complex operations may be broken down into simpler component operations, and a considerable reduction in the equipment necessary for this computer will result if the various simple operations are performed serially rather than simultaneously. As an example, the operation of multiplication can be broken down into a combination of additions, and if these additions are done in sequence, the same equipment can be employed repeatedly for addition, avoiding the duplicate equipment which would be needed to carry out many additions at the same time.

The result of such thoroughness in applying the serial principle to the reduction of equipment is greatly to lengthen the time required to solve any problem. There is, however, an immediate remedy for this, to be discussed in the next section.

2.6 ELECTRONIC DEVICES NEEDED FOR SPEED

To reduce materially the amount of equipment required for the solution of complex problems, it has been seen that a single quite versatile non-linear element, which we may designate as a computer, must be used to scan serially a large number of memory elements. To obtain solutions by this method in any reasonable time, it is obvious that most of the apparatus

must operate at very high speeds. The time spent at any one "network point" must be small, and the time required to transfer from one network point to another must likewise be small.

This dictates that the device employed must be electronic.

Ordinary vacuum tubes can generate and control signals which change rapidly within less than one microsecond. Vacuum tube circuits which will function to make or break a circuit within one tenth of a microsecond are already well known and require no elaborate design. On the other hand, electromagnetic relays must be specially designed if they are to operate reliably in 10,000 microseconds and no relay is at present known which will be better than this by more than a small factor. Consequently one may say that the vacuum tube is anywhere from 10,000 to 100,000 times faster than its electromechanical substitute.

If vacuum tubes are used, it is obvious that other elements of the device should be capable of responding to, and of producing, electrical signals although they may not be entirely electrical in their operation. The memory element which seems most suitable, and which will be described later, involves an acoustic delay line which has electrical input and output circuits. It is consequently desirable that the initial data of a problem should be fed into the machine electrically and that the results should be taken out of the machine by a device which operates electrically.

2.7 PROGRAM INFORMATION AND OTHER INPUTS

2.7.0 Classification of Input Information

Three categories of information must somehow be supplied in order to specify the problem which the machine has to solve. First of all, there must be a set of numbers on which calculations are to be made. Secondly, the machine must be provided in some way with directions as to what to do

with these numbers. This may be called the program information. Finally, many problems involve the use of particular mathematical functions which can be specified by a set of tabular values to be called a function table. The machine must be able to refer quickly to such a table and abstract from it the value of the function for any argument (within a specified range) which may occur in the course of calculation. It is of no value to achieve speed in the computing processes unless the three kinds of information which specify the problem can be imparted to the computing device with comparable rapidity.

2.7.1 Inputs to Analogy Machines

Let us consider briefly how these matters are handled in existing computing machines. In model devices, the model itself contains the information concerning the type of problem. The construction of this model may require a great deal of time and effort. In the network analyzer and similar analogy devices, the numerical information may be fed in by setting decade switches, while the specification of the operations to be carried out is inherent in the particular components selected and in the way in which these are interconnected. Arbitrary functional relationships are secured by the selection of components which have electrical characteristics such as to produce the desired relationships as closely as possible. In the differential analyzer, numerical information concerning initial conditions must be manually set by rotating certain shafts, and operating instructions are essentially contained in the mechanical interconnections which are established between shafts and various gears. Arbitrary functions are provided for by input tables with servo systems for following curves which specify the relationship between a function variable and an argument variable. The process of mechanically changing the problem which is set up

on a differential analyzer may require hours or even days. In order to overcome this, a differential analyzer has been constructed in which the interconnections are established by electrical controls which are in turn governed by a paper tape. This tape has holes punched into it and the positions in which these holes appear govern the operation of the machine. This arrangement accomplishes the purpose of considerably reducing the time required to set up the machine but obviously involves a considerable amount of equipment. It is understood that other methods of feeding in functional relationships are under development, but that servo mechanisms for following curves are still considered most practical.

2.7.2 Inputs to Arithmetic Calculators

Arithmetic as well as analogy machines are currently in use for the solution of complex problems. The way in which such machines receive numerical, program, and functional information is of particular interest here, in view of the conclusion (to be found in Section 2.8 following) that an electronic calculator must be arithmetic (or digital) in its operation.

First of all, consider a large class known as business machines. These machines store information on paper cards by means of holes punched in them in various positions. The operation of such machines is controlled chiefly by a plugboard or wiring unit with additional control from the cards themselves in certain cases. In their commercial forms these machines are capable of carrying out one detail of a calculation time after time as successive cards pass through the machine. When several different operations are to be done, each operation will in general require a modification of the plugboard or wiring unit. A great deal of the programming for a complex problem is therefore done by the operator in the process of taking batches of punched cards from one machine to another, arranging them in

CONFIDENTIAL

proper sequence and in changing the plugboards or wiring units of the various machines. Thus although a great deal of computational work may be handled in a mechanical fashion by such machines, it is obvious that involved calculations are, in no sense of the word, carried out automatically.

A calculator has been constructed (largely from components used in standard business machines) which will automatically carry out long and involved calculations. This calculator obtains its program information as well as much of its numerical data from a punched paper tape. Paper tapes are also used in this machine for the storage of arbitrary functions which may be needed in the solution of a problem.

Large computers have also been constructed utilizing electromagnetic relays both as switching devices and as memory elements. These devices too, have paper tapes for providing numerical information, program information and function tables.

The numerical results at the output of such a machine are usually stored in a form similar to the input; that is, either in punched cards or in paper tape. Either may be fed into some sort of automatic typewriter. In the present discussion, the speed at which printed records can be produced from such storage tapes or cards is not important.

For computers which utilize mechanical elements and relays as their essential components, the speed at which tape or cards may be read or punched is usually adequate, at least for providing numerical information and programming information. However, the storage of function tables on paper tapes often results in considerable loss of time because in hunting for a particular value a large amount of tape may have to be scanned. When function tables are stored on cards, an even greater limitation appears

CONFIDENTIAL

because existing card devices scan only in one direction. That is, the mechanism for feeding cards cannot be reversed.

2.7.3 Speed of Input and Control Devices

It has been seen that if complex calculations are to be done with a reasonable amount of equipment, then the calculations must be done serially. Further, these serial methods then require high-speed devices using electrical components to accomplish the desired results in a reasonable time. The programming information concerning a specific problem must therefore be available for rapid use by the electronic computing device. This can be assured by presetting a number of circuits, either by manual switches or by punched cards as in the ENIAC, but the time required to set up a machine for a given problem is then comparable to that required to set up the differential analyzer. If the program information is not set up statically in this way before the computation starts, it may instead be fed in serially as needed by the computer. The speed with which this may be done from a paper tape or card reading device is so far inferior to that demanded by an electronic computer that some other way must be found.

A satisfactory solution of this problem is obtained by the use of the same type of memory elements which are to serve for the storage of numerical information during the computations. As will be shown later, a memory element is available which is sufficiently inexpensive so as to justify its use for the storage of program information and function table information. It is therefore possible to store within such memory elements all of the data needed to specify and carry out the calculations for rather complex problems. All such information, then, may be set up initially and automatically in the electronic memory elements. These memory elements are

then sufficiently fast so that they will provide control signals to keep pace with the speed of the computer. Moreover, with the aid of electronic switching circuits, such memory elements can furnish functional values for any given arguments at speeds which do not impede the overall computation process.

Input and output devices which use a tape have inherent speed advantages over those which use cards. A card feed must take cards which are originally at rest and accelerate them so that they may be scanned by the reading mechanism. Thus even though the reading mechanism itself might be capable of scanning cards when run at a very high speed, there is a serious problem in the design of a feeding mechanism which will accelerate the cards without damaging them, yet at the same time pass them through the reading mechanism with proper tolerances on timing and location. A tape device could conceivably operate at high speed by running the tape continuously. As a matter of fact, existing tape devices use an intermittent motion whereby the tape is stepped along from one reading position to the next and is read only while it is at rest. There is no reason, however, why a tape cannot be scanned while in motion.

Paper is not the only material available for tape recording, and since the objective is high speed and reliable operation for the input and output devices as well as for the computer and memory elements, other materials should be considered. Various types of film such as moving picture film and special cellulose tapes are possibilities, but no other material seems as well suited for the purpose as a wire or tape of magnetic material, on which the information is recorded in the form of small magnetized regions. Such a wire or thin tape can store information much more compactly than it can be stored on conventional punched cards or

or paper tapes, and this information can easily be converted into a sequence of electrical signals. Moreover, such wire tape is strong enough to be run at high speeds, and it is not easily damaged by ordinary handling or by repeated use. Some of the alloys suitable for such application are naturally rust-proof and records kept in this form are therefore not only very compact but also very enduring. On the other hand, magnetic wire or tape records can easily be erased so that the same wire may be used over and over again if so desired. It is very likely that the cost of suitable tape or wire will also be less per unit of information stored than the cost of any other medium which might be considered.

A further discussion of magnetic recording will be found in Part 4. It is sufficient to note here that there already exist magnetic recording devices used for sound recording and reproducing which justify the statement that this method of handling the input and output of information from an electronic computing device is greatly to be preferred over any other existing method. The requirement that the input and output devices should be operated from electrical signals and generate electrical signals is nicely met by the use of this device.

2.8 ACCURACY REQUIRES DIGITAL COMPUTING

The foregoing sections have shown that the need for a general computing device capable of handling complex problems can be met without an unreasonable amount of equipment if the computing is done serially, and that this in turn forces us to seek electronic methods to recoup the loss in speed which such operation entails. So far nothing has been said concerning the accuracy to which such a machine should work. Model devices and most analogy methods such as the network analyzer can provide at most two or three significant decimal figures. In certain cases the differential

analyzer can work to four significant figure accuracy in some variables although the attainment of such accuracy is by no means easy. It is particularly difficult for an analogy device in which electronic tubes are used to achieve much accuracy if the various numerical quantities are represented by the amplitudes of currents or voltages in the circuit.

It is immediately obvious, however, that mathematical precision can be attained in a computing machine without corresponding precision in all of its parts provided that the machine operates arithmetically upon quantities which represent digits rather than attempting to insure accuracy by precise control of voltages or currents. One may note, for instance, that if voltages or currents can be controlled to within a few per cent, then a three digit decimal number might be specified by three different voltages, one of each representing the units digit, one the tens digit and one the hundreds digit of the entire number.

Since, however, the device now under discussion is to use commercial vacuum tubes and these are extremely variable, it is important to insure that this variability will not affect the accuracy of the computations. The best way of doing this is to go the extreme in which each tube is used purely as a switch. In this case any given tube is either nonconducting or conducting and no great accuracy of circuit components is necessary in order to insure reliable discrimination between the conducting and nonconducting states. Practical considerations concerning electronic tubes as circuit components require, therefore, not only a digital representation of numbers in the computing machine but the extreme digital system which may be called the binary system, wherein each tube has only the possibility of indicating a zero or a one. That the tubes are used in this way does not, of course, force adoption of the binary system of nota-

CONFIDENTIAL

tion at the input or at the output of the computer, nor does it require that the binary representation of numbers be used explicitly in any part of the machine. In the ENIAC the decado counters are made up of electronic tubes which are used always as described above, but there is no need to resort to the binary system of notation in order to work effectively with a machine which has such decado counters as components. There is then no requirement that a machine which utilizes tubes in this way must employ any particular number system in its operation, and a discussion of the merits of various number systems will not be undertaken until later in the report.

This section and the ones preceeding it have served to outline in an extremely general fashion the kind of machine which seems called for if one is to solve expeditiously complex problems such as those arising in advanced scientific theory and engineering design. To sum up briefly —

The machine must operate in a serial fashion to avoid large amounts of equipment; it must employ electronic components to compensate for the loss in speed which this serial operation brings with it; and it must use these electronic components as simple on-and-off devices in order that the required accuracy and reliability can be attained. Moreover, manual operations in preparing the machine for a particular problem can be avoided almost entirely since magnetic tape input devices can be employed to feed operating instructions in the same way that numbers are introduced into the machine.

2.9 THE ECONOMICS OF COMPUTING MACHINES

2.9.0. The Cost of Mathematical Solutions

As suggested in the introduction, reliable answers can often be found more quickly and cheaply by computation than by experiment, even though the experiment uses models. Results which are valuable if obtained quickly may become almost useless if a great deal of time elapses between the formulation of the problem and the attainment of the solution.

CONFIDENTIAL

Disregarding completely the high value which timeliness may possess, let us examine the "cost components" associated with mathematical solutions. These costs may be thought of as falling, for the most part, in the following general classifications: (a) Preparation of the problem in a mathematical form and logical symbolism appropriate to the machine, (b) Preparation of the machine itself in accordance with (a), (c) Actual computing, sorting, or other operations performed by the machine to achieve the solution, (d) Preparation of graphs or tables of the results in a form easily apprehended by those who are to use them.

For different types of machines, the costs in these categories must vary widely, and comparisons are more difficult because there is no fair way of deciding upon a "standard problem" or "typical problem" on which to base such comparisons. However, if attention is centered — as in this report — upon problems having the complexity associated with the integration of non-linear partial differential equations and integro-differential equations, then certain general conclusions are possible.

Even for problems of lesser complexity, it is evident that part (b) — preparation of the machine itself to carry out the solution — must be made as automatic as possible. Also, reduction in the labor and cost of (a) can be achieved if careful consideration is given to the structure and organization of the machine and its logical control. Obviously, too, the labor costs involved in (d) can be cut greatly if this can be carried out automatically. There is no point in pressing an extreme reduction in (c), the cost of the actual computing, if this is already low in comparison with the other three.

It seems best, nevertheless, to confine attention at first to the computer itself, partly because more definite and objective comparisons can

be made in this case, and partly because a discussion of this item will naturally bring to light some considerations pertinent to the other categories.

Speed of operation is relevant to the cost, but the speed which is of concern in this case is the speed per unit of computation. This is obvious since more computation can be done in the same time by duplication of equipment which may for instance be done by forsaking the principle of serial operation and carrying out various parts of the calculation simultaneously. We may then say that we are increasing the rate of computation by doing more units of computation at the same time; but to do this with the same type of equipment we must provide more units of equipment. There is no point then in comparing the overall speed of operation of two machines having different complexity but utilizing the same type of equipment for their unit operations. What is pertinent is an approximate comparison of machines which utilize different types of elements and this comparison to be significant must take into account the speed with which these elements operate.

Exclusive of the input and output devices, computing machines of the type here being discussed have three sorts of parts to be distinguished as memory units, switching and control units, and one or more arithmetic units which form the computer proper. Although the last of these may be considerably complicated because of the diverse functions which it must perform, the greatest amount of equipment will necessarily be involved in the memory, since in this there must be stored a large amount of information which is not being used at the moment but which may be called for on short notice. First consideration should therefore be given to the memory elements.

2.9.1 Relative Costs for Memory

A single relay may be used as an on-off or binary device. To do this one of its contacts may be used in a holding circuit so that, once the relay is closed, it remains closed until some other control opens the holding circuit. An electronic triode cannot be used in this way, but a pair of triodes, together with a few resistors and condensers, can be used in a circuit known as a flip-flop circuit which will function in much the same manner as the relay. Since two triodes for this purpose can be obtained within a single envelope or tube, one vacuum tube with its associated circuit components is roughly the equivalent of one relay for memory purposes. One flip-flop circuit of this type completely wired and ready to operate is to a first approximation just about as costly as one relay properly mounted and wired. As pointed out earlier the vacuum tube has an immense speed advantage over the relay. This advantage becomes important, however, only when the flip-flop is receiving information or when it is delivering its information to some other circuit and it is being cleared. During those intervals when the information is merely being held, the relay and the vacuum tube are more or less on a par. As regards maintenance and reliability, the relay may be preferable so long as this judged in hours and not in number of operations. In short, the vacuum tube has an advantage only when it is actively doing something.

When relays or vacuum tubes are used to represent numbers in the decimal system, inefficiencies are introduced since each of these devices is binary in nature, but these inefficiencies affect both the relay circuits and the vacuum tube circuits about equally and need not be considered here.

A point usually brought out in comparing vacuum tubes with relays is that relays can have multiple contacts, enabling a single relay to control or switch a number of circuits. Also, two or more coils can be wound

CONFIDENTIAL

on one relay, so that operation of this relay from several isolated circuits is possible. However, there is a price to be paid for these apparent advantages. Extra coils and contacts usually mean a reduced operating speed for a given type of relay. Also, in certain complicated switching and coding circuits, tubes have an advantage over relays which is not immediately obvious. In such circuits, there may exist unwanted current paths, described as "backcircuits", which must be eliminated. When tubes are used as switching elements, these backcircuits are prevented because the tubes will conduct current in only one direction. Relay contacts, however, lack this rectifying action, and additional relays must be used to open such circuits.

The advantage then which memory devices using vacuum tubes as flip-flops possesses over memory devices using relays is solely one of input and output speed. As devices for binary memory elements, the relative costs per unit operation cannot be compared without further information since the life of the relay is determined by the number of operations and the life of the vacuum tube is more or less independent of the number of operations but is measured rather in hours.

Let us suppose, for the purpose of making a comparison, that the life of a relay is 10^8 operations, and that the life of a vacuum tube used as a flip-flop is 10^8 seconds (about 3 years continuous 24-hour day use). (Relays and tubes exist for which these figures are similarly conservative.) It can then be seen that the flip-flop need operate an average of only once per second in order to equal the relay performance. Without resorting to elaborate circuit design, however, such a flip-flop can easily be made to operate at 10^5 times per second. As already pointed out, this speed is not used efficiently in a simple flip-flop memory system, but is thrown

CONFIDENTIAL

CONFIDENTIAL

away because, most of the time, the flip-flop is doing nothing. In a machine such as the ENIAC, where flip-flops are used in decade rings for memory purposes, only about one tube out of every 500 is "active" on the average during the actual computing operations. Allowing for a factor of 2 for times when tubes are on but computations are held up for input or output devices to function, we may say that each flip-flop is active about 100 times per second. Thus, even with the inefficiency attendant upon using the decade ring for decimal operation, the vacuum tube in its lifetime will perform 100 times as many operations as will the relay. (A further point which favors the tube is that, even though initial costs are about the same, tube replacement is quicker and cheaper than relay replacement.)

The factor of 100 for tube versus relay "cost per operation", rough and inexact though it may be, is meant only to illustrate what is believed to be a fruitful way of approaching the topic of the economics of computing machines. Whether this factor is correct for the example cited is not of any great importance. What is important is to note that since vacuum tubes are capable of operating with pulses at rates in excess of 10^6 per second, the flip-flop type of memory element is nowhere near the point of making good use of electronic tubes. A much more efficient use of vacuum tubes is possible in the serial memory device which is to be described in Part 3. A device of this sort capable of storing 1,000 pulses or 1,000 binary digits will require no more than ten tubes. This efficiency is obtained as a result of the fact that the 1,000 binary digits are to be fed in and out of this device in serial order. As the pulses are fed in from some other source, they must pass through certain shaping and timing circuits and are then introduced into a delay line. The pulses which emerge

CONFIDENTIAL

from this delay line have already been "remembered" and for a period of time equal to the delay which this line has. Now if this same pulse pattern is to be stored for a longer period of time, it is fed into the delay line once again, and may be recycled indefinitely. A unit of this sort which can store 1,000 pulses with about 10 vacuum tubes may cost approximately the same as 20 vacuum tubes and their associated circuits so that we may say that each binary element may be stored indefinitely for $\frac{1}{50}$ of the cost of storing this information in a flip-flop type of memory. For the cost per operation, the number of operations per second must be estimated, since the lifetime of such a device depends on the total time of operation. It is easy to make such a device work at a million pulses per second; and if each such device in a computer were active only $\frac{1}{1,000}$ of the time, as was estimated for the flip-flop in a computer like the ENIAC, then we would arrive at 1,000 operations per second as an average figure. We may now compare the cost per operation of the delay line memory with that for the flip-flop type of memory: From the figures just given, we estimate that the delay line memory is about 500 times cheaper on this basis. However, this comparison is not necessarily correct since the tubes used in the delay line memory may actually work a greater fraction of the time than those in the flip-flop type of memory. It is to be noticed that we must distinguish between the operations which the tube may go through in the course of regenerating the pulse pattern and the useful operations from the mathematical point of view wherein the number is transferred from one place to another and otherwise manipulated. Taking figures which later sections show to be reasonable, we might use about 100 such memory devices in a single machine and obtain 100 or so pulses from some one of these for each cycle of operation. From this it is seen that the estimate that one memory element is

used only $\frac{1}{1,000}$ of the time is proper. This may be seen directly by noting that this memory capacity amounts to the storage of somewhere between 1,000 and 2,000 numbers (each having 8 or 10 decimal digits). At any one time only 1 or 2 of these can be utilized in the computing mechanism.

The cost of this delay-line type of memory when used for the storage of program information should also be considered. If we ignore the speed factor entirely and treat this memory device solely as a static storage element, it is much better than relay devices and is about on a par with manually set switches.

It is therefore entirely reasonable on an economic basis to utilize these memory elements for the storage of program information and function table values thinking of them merely as static storage elements, and without taking into account the ease of setting them up and the speed with which they operate and their number of operations per lifetime.

CONFIDENTIAL

2.9.2 Comparison of Input and Output Devices.

A somewhat similar discussion can be given for the various lower speed devices for handling and recording data in permanent form. Cards which can receive 80 or 90 decimal digits can be fed through reading equipment at the rate of 1 or 2 per second. Paper tapes may read one decimal digit per tape position and 3 to 5 positions every second, or wider tapes may be used with somewhat increased speed. Each channel on such a tape has an "input frequency" of 5 pulses per second. By the use of 5 or 6 channels (or up to 30 as has been done) the effective input frequency is considerably increased. Since these are multiple channels this increases the equipment. A business machine using punch cards, fed at the rate of 2 per second and having 10 or 12 hole positions in each column, could be used* as a device having 80 or 90 input channels each of about 25 pulses per second input frequency. The effective input frequency of 2,000 pulses per second is considerably higher than that obtained from most tape devices, but this high rate is achieved by the large number of multiple channels which again means more equipment.

In contrast to these figures, consider those for a magnetic wire or tape recorder. There is a possibility that devices can be developed to handle 30,000 pulses per second through a single channel. However, a conservative estimate, based on the performance of existing commercial sound recorders which use wire, is 5,000 pulses per second. This rate through a single channel is two and one-half times greater than the rate which card machines achieve through 80 or 90 channels, and about 200 times greater than the rate for a single channel of a card machine. Although the relative cost per channel for these devices cannot be estimated at this time,

* Normally used with only one hole per column, the information on a card is equivalent to only 320 pulses, so that two cards per second is equivalent to 640 pulses per second.

it is unlikely that a magnetic recorder channel would cost anything like 200 times as much as a card reading or punching channel. Accordingly, it is inferred that the relative cost per unit operation is probably much lower for the wire device.

2.9.3 Graphical versus Tabular Outputs

The reader is referred to the second paragraph of Section 2.9.0, where the costs associated with mathematical solutions were classified. It now appears that, through the use of a new high speed memory device and magnetic wire or tape recorders, a material reduction can be made in items (b) and (c). In fact, this reduction can conceivably make items (a) and (d) overshadow (b) and (c). Reduction of (a) and (d) is important to maintain a reasonable balance. Item (d) will be considered first.

Automatic printing of tabular material from magnetic wire or tape records is obviously necessary. Fortunately, there is no inherent difficulty in making a device to do this. A typewriter, operating serially to print one character at a time, is in keeping with the serial operation principle advocated for the computing machine itself.

When the results of a computation are extensive tables of numbers to be distributed to many potential users, photographic reproduction of the tabulations made by such an automatic printer will be used. All the necessary work to put such tables in a useful form can be done automatically. All differencing, smoothing, inverse interpolation, sorting and arranging can be carried out by the computer. No reason remains for issuing tables which are not in a form convenient to use.

The preparation of extensive numerical tables is a possible use for a computing machine, but by no means the only, or even the chief, use.

CONFIDENTIAL

For many design problems, a number of solutions may be carried out for just one purpose -- to find that one solution which best satisfies some particular criterion. Several parameters may be at the disposal of the designer. He may know roughly what values to give them, or he may not, but in any case he wishes to find "satisfactory" values. He may be interested in the allowable tolerances in some parameter, that is, how sensitive some "result" is to possible variations in the parameter.

The pertinent information for design problems like these can be presented graphically. Maxima, minima, inflexion points, and all sorts of qualitative and semi-quantitative relationships between several variables can be taken in visually from a family of curves. Even after long and careful scanning of tabular material, some of the features which are graphically "self-evident" may be missed.

For problems of this sort it should not be necessary to wait for tabulations to be printed, and then to wait even longer while a great deal of labor is spent in plotting the tabular values and in drawing families of curves. For it is, in fact, easier to make an electrical computing machine operate a cathode ray tube than it is to make it operate a typewriter. A high-speed computing machine can, in many instances, compute results as fast as they can be taken in by the eye when "plotted" on the screen of a cathode ray tube. If, for more complicated problems, the results are not turned out rapidly enough for immediate transfer to the visual output, these results can be stored in the memory and "run off" when completed.

The usefulness of a "visual output" is not confined to design problems of the type cited above. In the preparation of numerical tables one does not always know beforehand what tabular intervals must be used to

keep certain differences small, or what the range of variation of the quantity calculated is going to be, or whether certain numerical processes will converge nicely. It is desirable to have some way of investigating these things quickly and easily. As before, no great accuracy is needed to judge these matters. A visual output can show immediately when a particular quantity has gone "off scale". The smoothness of a tabulation can be checked by having differences of a specified order transferred to the screen.

2.9.4 Preparation of the Problem

The cost of steps (b), (c) and (d), as classified in Section 2.9.0, have now been examined, and devices have been proposed which can be expected to reduce these costs considerably. All three steps are to be carried out automatically from directions which are stored magnetically on a wire or tape. To prepare input data in this form, a machine something like a typewriter will be provided. If step (a), the preparation of the problem in a form and symbolism appropriate to the machine, has already been done, then an operator with no mathematical knowledge can transcribe the problem data to the magnetic wire or tape.

The cost of step (a) remains to be considered. This cost may easily be the largest of the four, and hence reduction of this item is extremely worthwhile. How may this be done?

It may be done by giving careful attention to the organization and control of the calculating machine, and to the code which operates the controls. Only a few basic orders must be employed, so that constant reference to a dictionary is avoided. But these basic orders must be chosen so that, by simple and easily learned combinations, a great many frequently

desired operations can be carried out. An order to add two numbers might be considered basic. Should separate orders be used to transfer the desired numbers to the computer, or should one order combine the directions to add with the specification of the position in the memory where each number is to be found? Decisions as to how arithmetic operations are to be specified by an order code can probably affect the cost of step (a) only slightly.

The cost of step (a) is, however, immensely sensitive to the way in which repetitive and interative processes are ordered. Essentially, the machine must be made to perform "substitutions". For instance, the same sequence of arithmetic processes must be carried out over and over again, but each time on a new set of numbers. To repeat such arithmetic orders each time they are to be used would be wasteful of memory capacity as well as wasteful of coding time, and is obviously unnecessary. As another example, interpolation in a function table may be called for at various times during a calculation. The directions for carrying out this work need be inserted in the memory only once, and the basic order code must provide for finding these instructions each time that they are needed, and for returning to different parts of the problem orders after each use.

(Another type of substitution is also desirable in using function tables. Hunting through the table for a certain argument can be avoided by having the argument specify the place in the memory at which the desired function value is to be found. But to do this, the order code must be made with provision for substituting a number (obtained from that part of the memory which holds the argument) into an order, so as to specify where the function is to be found.)

CONFIDENTIAL

It is by the use of "subsidiary chains" of orders, to be called into use from time to time, as they are needed, by a "higher" set of orders, that a computational routine can be compactly represented. What is more, this corresponds to the way in which mathematical processes are most easily and naturally thought about. The rule for interpolation is not written down anew every time it must be used, but is regarded as a "subsidiary routine" already known to the computer, to be used when needed.

This compounding of orders at different levels (sometimes referred to as a hierarchy of control) constitutes the most important strategy for easily organizing and writing down the instructions for a large computation. It is therefore important that the machine and its method of control be planned so as to make effective use of this principle.

CONFIDENTIAL

PART 3 THE DELAY LINE MEMORY DEVICE

3.0 THE DELAY LINE MEMORY DEVICE

A high speed memory element which is simple enough and cheap enough so that a large memory capacity can be obtained with a reasonable amount of equipment is an essential component of the EDVAC. In this section a description will be given of a memory device which satisfies these requirements and which seems the best available device for use in the EDVAC.

The memory device uses a delay line as a serial storage device and regenerates the signal pattern so that virtually unlimited storage times may be obtained. It is essential that the signal patterns to be stored are digital in nature, but this is already one of the requirements which must be placed on the computing system in order to insure accuracy. Also, it is necessary that this device be kept in synchronism with a standard clock or oscillator, but this characteristic is desirable for use in a computing machine.

3.1 OPERATION OF THE DELAY LINE MEMORY

The elements of such a memory device are shown in the block diagram of Fig. 1. C is an oscillator which serves as a clock. It emits continuously a train of pulses which may for the sake of illustration be considered as similar to those illustrated in Fig. 2c. L is a transmission line with characteristics such that pulses which are fed into it at one end are transmitted and issue from the other end after a delay time T. The output pulse from L need not be exactly like the input pulse as will be seen later. In general such an output pulse may be passed through an amplifier and may then be used as a signal upon grid of a gate tube G. The second control grid of the gate tube G receives the continuous train of clock pulses from C. G is a coincidence device, and a pulse appears in its output only when G receives simulta-

Simplified Block Diagram for Delay Line Memory Device

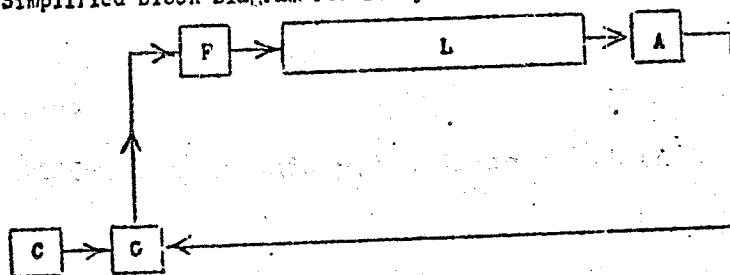


Fig. 1

- F Pulse former
- L Delay Line
- A Amplifier
- G Gate (coincidence tube)
- C "Clock": source of timed pulses

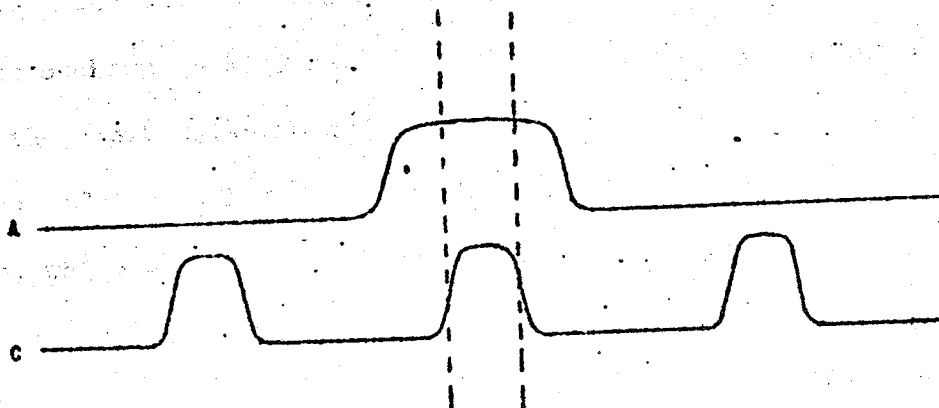


Fig. 2

This figure illustrates the approximate shape which pulses fed to G from A and C (in Fig. 1) would have.

neously pulses from A and C. The output pulse from G is then fed back into L, preferably through a forming circuit, F, which gives the input pulses to L a shape and amplitude which is optimum with respect to the characteristics of L.

Let us now assume that in some way a single pulse is introduced into the circuit FLAG and that this pulse is so timed that it coincides with a clock pulse at G. In consequence, a new pulse will be set up through the same circuit and after a time T this pulse will arrive at gate G. If T is made a multiple of the period of the clock C, a new pulse will again be caused to circulate through L. This repetition will continue so long as two necessary conditions are maintained. One of these conditions is, of course, that the time T shall sufficiently approximate some multiple N of the clock period, and the other is that the total gain around the circuit FLAG shall be unity. The first condition is not a severe one because the pulses which must coincide in G are not infinitely narrow and can be made broad enough so that slight variations will not interfere with the coincidence. As a matter of fact, the pulses fed from A to G will very likely be broader than those from C and can be deliberately broadened if necessary so as to resemble those shown in Fig. 2A. In this case the output pulses from G will resemble the clock pulses C and their timing accuracy will be determined by the accuracy of the pulses C even though T varies slightly. The second condition, which has to do with the gain around the circuit, is easily provided for by well known methods of gain control and the limiting action of amplifiers.

It is seen then that if a single pulse is introduced into this circuit at the proper time, there will be a new pulse sent through the delay line L at intervals which are accurately N times the clock period. But

CONFIDENTIAL

more than one pulse may be used in this circuit. In fact, as many as N pulses may be introduced, and each will separately be perpetuated in this way.

In practice, the transmission line L can be an electrical transmission line if N is a small number like 4 or 5, but for larger values of N attenuation and phase distortion in electrical lines makes it impossible to use them for this purpose. However, the transmission line need not be purely electrical if it can be provided with electrical input and output devices. The best device now known for L when N is large consists of an acoustic transmission line which uses piezo-electric crystals for its input and output. When the input crystal is electrically pulsed, a mechanical or sonic pulse is transmitted through an elastic medium. Traveling at the velocity of sound in this medium, this pulse arrives later at the other crystal and causes it to generate an electrical output signal. Solid materials for the acoustic line have the disadvantage that there are several modes of propagation each with a different velocity. This can be avoided by the use of a liquid so that only longitudinal waves can be propagated. The liquid which best nearly fulfills the requirements for this purpose is mercury since this has a low velocity of propagation and yet, because of its density, has an acoustic impedance comparable to that of quartz crystals thus preventing disturbing reflections. Acoustic delay lines using mercury were developed at the Moore School as accurate timing devices for other applications in 1942. In 1944, during work on the ENIAC, the regenerative circuit which allows this delay device to be used as a memory unit was disclosed and successful preliminary experiments were performed. The method of insuring accurate timing by gating pulses was one which had already been utilized in the ENIAC.

It is obvious that the addition of a few gate tubes to the circuit of Fig. 1 enables this memory device to be cleared of pulses and to be fed new pulses. Also the circuit may be made to transmit to other devices a pulse pattern similar to the one which it holds and so long as its own circuit is not cleared it can repeat such transmissions as often as desired. It is interesting to note that this memory device can if necessary perform all three of these functions at once. That is, it may transmit the signals it has been storing, clear itself of these signals, and simultaneously receive and store a new set of pulses.

3.2 EQUIPMENT AND STORAGE CAPACITY

From experimental measurements on acoustic delay lines it is known that approximately 3 stages of amplification (for pulses 1 microsecond apart) will be necessary to offset the attenuation occurring between the electrical input and the electrical output (most of this attenuation occurs in the input and output crystals and therefore is independent of the length of the delay line). In a machine which uses a number of memory elements only one clock circuit C need be provided so this part is not to be considered when counting the equipment needed for each memory element. The various gate tubes and the forming circuit F must, of course, be provided for each delay line. It is estimated that the number of tubes required for these and for the amplifier will total about 10.

Experiments have shown it to be practical to utilize this memory device so as to store 1,000 pulses occurring one microsecond apart. This device is therefore capable of storing 1,000 binary digits and yet uses only 1 vacuum tube for each 100 digits stored. As a matter of fact, among the 10 tubes included in the above estimate are several which are used for switching purposes. If such switching means were similarly included in estimates for

relay or flip-flop memory systems, then one relay or tube per binary digit would no longer be sufficient. Considering this, the comparison is even more favorable toward the acoustic delay line memory unit.

Electronic beam tubes utilizing a memory depending upon the storage of electric charges may be able to do better than this, but at present the development necessary to make such a system practical is hard to estimate and might be very considerable. These two types of memory are the only ones which at present seem capable of operating at the speed of 1 pulse per microsecond. Cheaper memories can be obtained by the use of magnetic devices, and magnetic wire records are an example of this. At present, however, these are not capable of operating at the speed just mentioned and therefore are to be used in the EDVAC at a lower speed for permanent or long-interval storage.

3.3 TERMINOLOGY FOR EDVAC MEMORY

In order to facilitate discussion of the application of the delay line memory device to the EDVAC, the following terminology will be introduced. A delay line memory unit will be called a tank. Each pulse pattern representing a number or an order will be called a word*. To simplify timing problems, the same time must be allowed for every word. Hence, to use the tanks efficiently, all words, whether numbers or orders, should be approximately the same length. A tank designed for the storage of many words will be called a long tank, and one which is designed to store only one word will be called a short tank. To permit time for switching operations, the words in a long tank should be separated by two or three microseconds. The time from the beginning of one word to the beginning of the next will be called a minor cycle.

*These terms were suggested by von Neumann

It seems desirable to have only one length for long tanks, and to call the corresponding time a major cycle*. A short tank repeats its word every minor cycle, and a long tank repeats all of its words once each major cycle.

3.4 SHORT VERSUS LONG TANKS

It would be ideal to have every word, no matter where stored, available during any minor cycle. This could be done by using only short tanks. There are two chief disadvantages to this solution, however. Long tanks involve no more electronic equipment than short tanks, and therefore cost little more than short tanks, yet each long tank is capable of storing 30 or 40 words instead of one. Hence, large memory capacity is provided most cheaply by long tanks.

A second disadvantage of using all short tanks appears when switching problems are considered. Suppose that the desired memory capacity could be provided either by M long tanks each containing N words, or by MN short tanks. The short tanks then require an MN -way switch, while for the long tanks an M -way switch suffices. The N -way selection is then a time selection, effected by a properly timed gate in the trunk-line common to all tanks. Also, the output of each tank can drive only a certain limited electrical capacity, and therefore, when more tanks are to feed into the same trunk, they must be coupled to this trunk in groups with additional driver tubes for each group. This last fact is more serious than the extra switching requirements, since the complication of the switch is more or less balanced by the fact that a counter circuit must be used with the long tanks to determine when the trunk-line gate opens to select the desired word from a long tank.

*This term was suggested by von Neumann

CONFIDENTIAL

A compromise is certainly possible, wherein both long tanks and short tanks are used. The long tanks can then provide most of the desired capacity, and short tanks can be used in such a way as to reduce the average waiting time. In Part 7, specific numbers of long and short tanks will be suggested.

PART 4 MAGNETIC INPUT AND OUTPUT DEVICES

4.0 GENERAL FEATURES

The discussions of sections 2.7.3 and 2.9.2 have already indicated that devices which employ magnetic wire or tape are desirable for the input and output mechanism of a high speed computer. Magnetic recording devices have been known for a long time, and recently commercial sound recorders using magnetic wire have been used for many purposes. The application of such methods to the recording of numerical data in digital form requires a few modifications and developments, but presents no formidable difficulties.

In some respects, the requirements for digital recording are not as strict as those for sound recording. Linearity of response to signals of various amplitudes is important in sound recording, but not in pulse recording. All that is required is that a pulse be recognizable as a pulse--not that it have some precise amplitude. This is similar, of course, to the use of vacuum tubes as simple on-off devices, so that their linearity is not important. Likewise, frequency distortion is allowable so long as it does not interfere with pulse recognition. And finally, noise levels which would be intolerable for sound reproduction may be permissible for pulse recording. Reliable discrimination of pulses from noise can be assured with a signal-to-noise-level ratio of four or five.

CONFIDENTIAL

On the other hand, the problem of starting and stopping the wire or tape is of no importance in sound recording, but is very important for the use proposed here. Much tape (and time) would be wasted if it were not possible to bring the tape from rest up to recording speed and stop it again quickly after a given piece of information has been recorded. The necessary control mechanism for this purpose will be discussed in Section 4.6.

In passing from magnetic recording pulse rates to microsecond pulse rates, and vice versa, problems of timing and synchronism arise. These are nicely handled by methods to be discussed in the following sections.

4.1 RECORDING FROM HIGH SPEED MEMORY

The rate at which pulses may be recorded magnetically on wire is much lower than one per microsecond. Moreover it is desirable that the process of reading from wire records into acoustic memory units shall not require extremely accurate control of the wire speed. These two conditions are well taken care of by the method now to be explained.

Suppose that a pulse pattern which is to be transferred to a wire is at present stored in a delay line memory for which the minor cycle is 32 microseconds. This memory utilizes a short acoustic line of 31 microseconds delay and an electrical line of 1 microsecond delay. (In this explanation, other delays will be considered negligible.) Ordinarily, the signals to be regenerated will enter the electrical line and then, after one microsecond, pass into the acoustic line. The pulse pattern will thus be repeated unchanged each minor cycle. A gate tube has one of its controls supplied with a gating pulse, P , occurring at the first pulse time of each minor cycle, and its other control is fed from the memory system just described.

It will be convenient to assume that, of the 32 possible pulse positions (or times) in a minor cycle, the first and last are always blank, and the pattern or word occupies the other thirty. Then as long as the one microsecond line is kept in the memory system, no memory pulse coincides with P, and there is no output from the gate tube.

At the time that it is desired to begin recording, the wire may be at rest, in which case it is accelerated quickly by a servo device. When the wire speed has been brought within the proper range, the one microsecond delay is electronically switched out of the memory circuit, with the result that the pulse pattern in the memory circuit advances by one microsecond during each minor cycle. The pulses "pass in review" before the gating tube where F is applied. The output from this gate tube is fed to the recording magnet, and, at the rate of one pulse per minor cycle, the entire pulse pattern from the tank will be transferred to the wire in exactly the same temporal sequence. This is accomplished within 32 minor cycles. If desired, as each pulse is recorded on the wire, the corresponding pulse may be deleted from the tank, but this is not necessary in the particular arrangement just described.

The method which has just been outlined requires the magnetic wire recorder to accept pulses at the rate of one every minor cycle, or at a rate of about 30,000 pulses per second. Information from other laboratories indicates that such a device might be within the range of possibility. Such speed is not essential, however, as a modification of the plan described above allows recording and reading to be done at lower pulse rates. After experimentation with wire recorders, it will be possible to fix a pulse rate which can be reliably secured without protracted development. As a lower limit to this pulse rate, it is known from the characteristics

of commercial sound recorders using magnetic wire that a pulse rate of 5,000 per second is easily obtained.

To obtain approximately this pulse rate from a short memory tank, it is necessary only to provide for five "normal" minor cycles after each minor cycle which advances or retards the pulse pattern. The P pulses would then be gated with a minimum spacing of six minor cycles, or 192 microseconds, so that the pulse rate to the recorder would be just slightly greater than 5000 per second.

4.2 READING INTO THE HIGH SPEED MEMORY

The process just described is capable of reversal, so as to read from a slow pulse rate system into a fast pulse rate system, provided the slow pulse rate system is also accurately timed. But it is not practical to control the speed and position of the wire accurately enough to synchronize with a microsecond clock. A synchronizing circuit, similar to some used in the ENIAC, avoids this problem. Pulses from the wire are fed into a flip-flop which acts as a temporary storage device. Once each minor cycle, a pulse is applied to a gate controlled by this flip-flop. If the flip-flop is set, a pulse is inserted into the memory tank, and the flip-flop is reset to await the next pulse from the wire. In the minor cycle in which this occurs, an extra delay of 1 microsecond is introduced into the tank regeneration circuit, so that all pulses previously stored in the tank are "retarded one place".

However, possible pulse positions which are blank must be allowed for. If the wire speed were accurately controlled, the absence of a pulse could be inferred by timing and a corresponding blank introduced into the memory tank by using the 1 microsecond delay. Numbers represented in the excess-three binary decimal code always have at least one pulse present in

each group of four. For certain order codes, longer stretches of blanks are probable. To correctly sense the proper number of blanks in such symbols, it would probably be necessary to control the tape speed to within one or two per cent.

This difficulty can be avoided by the simple expedient of actually recording blanks on the wire in some way. The method which at present appears simplest is to use pulses of opposite sign (reversed magnetization) when blanks are to be indicated. This also affords an easy way of checking the operation of the recorders and readers, and of obtaining signals for a rough control of wire speed, if such control is desired.

4.3 REVERSIBILITY OF WIRE DIRECTION

In the method just explained for recording and reading, the pulses in the delay line memory unit were "advanced" in position by temporarily removing a 1 microsecond delay line. By this means, the entire pulse pattern in the memory unit was made to pass in review before the gate where P was applied.

It is easily seen that the pulses may be made to pass in review in reversed order by inserting an extra one microsecond delay, instead of removing one, so that the total delay becomes 33 microseconds. If, using this plan, the wire is run in the same direction as before, the pulse pattern on the wire will be reversed. But if the wire direction is reversed, then the original sequence of the pulse pattern will be restored. (When the wire direction is reversed, electronic switches interchange the connections to the recording and reading magnets, since their functions are also reversed. That is, the reading magnet then serves as the recording magnet, and vice versa. The same controls which accomplish this interchange will also determine whether the delay is to be 31 or 33 microseconds.)

Thus, exactly the same pulse pattern can be put on the wire from the short tank, regardless of the direction in which the wire is running. A similar technique can be applied in the reading of magnetic pulses from the wire. The ability to record and read correctly independent of the wire direction is important. For one thing, it saves the time of rewinding, and makes a long wire practically the equivalent of a long loop, without the disadvantages of handling long loops.

4.4 COMPACTNESS OF WIRE OR TAPE RECORDS

In previous sections, emphasis has been placed upon the speed with which information could be read or recorded by magnetic wire or tape devices. A further important advantage in having records in magnetic form is that this medium is lighter and less bulky per unit of information stored than any other medium. The estimates which follow are based on using a wire four thousandths of an inch in diameter which is approximately what is used in commercial recorders.

It is expected that at least 100 pulses per inch could be placed on such a wire. This figure depends mainly on the construction of the pole pieces and associated structure, and not on wire speed. (The wire speed must then be directly related to the pulse rate, and for 5000 pulses per second the wire speed must be at least 50 inches per second. Speeds of 7 ft. per second are common, and much higher speeds are possible.)

The information recorded decimally on an 80 column card is equivalent to about 300 pulses, or about 3 inches of wire. Spools about 4 inches in diameter and one inch wide are used commercially to hold about 12,000 ft. of wire, which is roughly the equivalent of 50,000 completely filled cards. This wire weighs about 1/2 lb, and the weight of the spool is probably of

the same order of magnitude. For various reasons, the cards may have to be used inefficiently (only half filled, say) while this is not true of the wire. Hence cases can arise where this spool of wire is the equivalent of 100,000 cards. The weight and bulk of this many cards is very large compared to the weight and bulk of the wire and spool.

This is an important consideration for storage purposes, but it is also important in the way it affects calculating and sorting operations. Handling 50,000 or 100,000 cards, and seeing that they are fed into and taken out of a card machine, is to be compared to placing several spools of wire in magnetic devices which thereafter operate automatically. Cards may easily get shuffled out of their proper order, while the information on the wire is not subject to such accidents.

4.5 SORTING AND COLLATING

4.5.0 The Need for Sorting and Collating

Sorting and collating are not always thought of as computing operations. Nevertheless, there are numerous ways in which a need for these operations may arise in the course of computing problems. Consequently, it is important to consider how sorting and collating can be done efficiently and at high speed.

It is a common notion that machines which employ cards are peculiarly suited to this work--and that tape machines are inherently at a disadvantage. This belief has been encouraged by the fact that commercially available tape devices are much inferior to card machines for sorting and collating.

If during a computing problem certain sets of numbers which are held in the high-speed memory must be sorted or collated, it is obvious that this should be done within the machine if possible. There is, in fact, no obstacle to this, and it is relatively simple to provide the controls and instruction code to take care of such situations. It is estimated that the sorting and collating speeds attainable within the computer will be consider-

ably higher than those now achieved by card machines.

Given that the machine is thus equipped to do "internal sorting", and "internal collating", it is tempting to think that the general problem has been solved by the expedient of transferring all tape information to the machine, doing the job within the high-speed electronic part, and then putting the results back onto the tape. This "solution" is, in fact, quite inadequate for large sorting jobs, since the memory capacity of the electronic machine is necessarily limited, while the number of cards (representing the memory) which may be processed by a card sorter is essentially unlimited. The magnetic wire or tape is the corresponding unlimited memory for the electronic machine, and sorting and collating processes must be devised which will utilize this memory without being restricted in any way by the limitations of the internal high-speed memory.

4.5.1 Tape Versus Card Sorting

The essence of sorting and collating is to take information which is arranged in one order (which may be random) and rearrange it in some other order. When the information is punched into cards, the information is "moved" from one place to another by moving the card which carries that information. When paper tapes are used, this method is no longer feasible, and the information itself must be moved by perforating the same information into a new tape. The only inherent disadvantage of this process is that the old tape may be of no use, so that in each such transfer of information, paper is wasted. A large number of transfers may have to be made before the sorting process is completed, so that the wastage is by no means unimportant. Magnetic tape or wire, on the other hand, immediately overcomes this problem since it can be erased and used for new storage as often as may be desired.

CONFIDENTIAL

The fact that sorting is done more slowly with paper tape than with cards has nothing to do with the fact that one medium is "continuous" while the other is "discrete". The problem is solely that of how fast the information can be "moved". It so happens that commercial devices using paper tapes move the tape intermittently, reading it while at rest, and then advancing it to the next reading position. Also, such tapes usually have only five or six channels, while cards have 80 or 90 columns. As brought out in Section 2.2.2, paper tape devices of this sort have an effective rate of one thousand or more pulses per second. The great speed advantage of the card device over the paper tape device comes not from any inherent property of cards as opposed to tapes but from the fact that (a) card machines have many more channels, and (b) card machines have been designed with speed in mind, while the paper tape machines have been designed to work with other equipment which is also slow -- for instance, a serial typewriter.

(Much higher speeds are known to be possible in devices which read paper tape photoelectrically. It is likely that perforating speeds could be increased, but not to the same extent. A single channel of a card punches holes at an approximate rate of 30 per second, while a single channel of a paper tape machine ordinarily punches 5 holes per second.)

Since a magnetic wire or tape device can, through a single channel, read or record at least 5000 pulses per second, information can certainly be moved from one place to another at least several times faster by these means than by card machines, in spite of the fact that the latter use 80 or 90 channels. Since this is so, sorting and collating should be done with magnetic devices rather than with card machines.

CONFIDENTIAL

4.5.2 A Binary Sorting Device

The type of equipment which is needed for sorting and collating will now be described. The most elementary sorting process is one in which a single class of data is separated into two classes. This may be called binary sorting. The corresponding binary collating operation is that in which two ordered classes are meshed together to form a single ordered class. All more complicated cases of sorting or collating may be performed by a sequence of such binary processes. (It may be noted that card machines usually have binary collators, but use decimal, or even twelve or thirteen-pocket sorters.) At least three tapes or wires are needed in order to carry out these binary processes easily, without recourse to an extensive additional memory device.

Let it be supposed that a single magnetic wire contains a series of words which represent numbers, and that those numbers are in random order. It is desired to arrange them in "natural order", that is, in monotonic sequence. If each number is specified in terms of its binary digits, then the first step is to sort on the units digit. That is, as successive numbers are read from the original wire, all odd numbers are to be recorded on one wire, and all even numbers on another. This is accomplished by passing the pulse trains from the reader through an electronic discriminator circuit which then gates each word to the proper recording station, the proper gate being selected by the presence or absence of a "one" in the units place of that word. At the end of this operation, the original wire will have been erased (unless there is some reason for preserving the information in its original order for other work). Further binary sortings are now required (on successive digit positions), and two "clear" wires must be available. This could be arranged by transferring all the odd

CONFIDENTIAL

numbers to the even wire, or vice versa, but such transfer is wasteful of time and can be avoided by providing four magnetic wires instead of three. It will therefore be assumed that four wires are used.

The second binary sorting process can then be begun by reading from the wire having odd numbers, and discriminating on the second rather than the units digit to determine which of the two free tapes will receive each word. After all of the odd numbers have been classified in this way, the even numbers are likewise classified by their second digits. They are recorded on the same two wires, but since they follow the odd numbers, the result of the first sorting has not been lost. At the end of this sorting process, there will again be two wires "clear", and sorting on the third digit can be carried out in a similar manner. Words consisting of N binary digits can therefore be arranged in monotonic sequence by N binary sortings, and during each such sorting these words are transferred from one wire to another at the rate of 5000 or more pulses per second.

It is in operations such as that just described that the practical necessity of being able to start and stop the wire or tape is keenly felt. Were every wire to run continuously (except for reversal at the end of each binary sort) then each binary sort would result in doubling the length of wire needed to hold the data. To waste wire in this way is undesirable, but of even greater importance is the fact that the time required for completing each sort is proportional to the length of wire which has to be fed through, whether that wire is blank or not. Furthermore, in cases where the original data occupied an entire spool of wire, then even more time would be wasted because many spools of wire would have to be handled by the operator before the job was finished. However, the serve devices to be discussed in section 4.6 overcome such difficulties. Whichever wire is not being used at

CONFIDENTIAL

any moment is at rest, and no blank spaces occur. The total length of wire occupied by the data after N binary sortings is equal to the length of wire occupied by the data in its original form. Each binary sorting therefore takes the same length of time as any other.

4.5.3 Sorting Speed

It is now possible to make some remarks concerning the speed with which sorting may be done by magnetic wire methods and compare this with card speeds. As a conservative estimate, take 5000 pulses per second for the rate at which information is transferred by magnetic devices, and assume that this figure already allows for the acceleration and deceleration times. (The justification for this is in Section 4.6) Further assume that a "full card" carries 80 decimal digits, that each digit is equivalent to four binary pulses, and that a card sorting machine will handle eight cards per second. Thus information is passed through a card sorter at about 2500 pulses per second when the cards are completely filled. The binary sorting device using magnetic wire passes information at twice this speed, but four binary sortings are necessary to equal one decimal sorting. Hence in this case the single-channel magnetic sorter is only half as fast as the 80-channel card sorter.

This comparison is, however, biased in several ways. In the first place, a very conservative figure has been used for the magnetic device, while the card device has been favored by assuming that the cards were completely filled. If the cards were only half filled, and the magnetic device were operable at 10,000 pulses per second, then the magnetic wire sorter would be twice as fast as the card sorter. Further, this comparison has neglected the fact that the magnetic wire device is fully automatic in operation, whereas the volume of data which can be placed on one spool of

wire might require 100,000 cards, so that an operator would be kept busy several hours attending to the card sorter. Furthermore, if sorting is to be carried out on more than one card column, or decimal digit, a great deal more card handling is necessary. This not only takes time, but introduces the possibility of errors on the part of the operator. Finally, if the required sorting is in fact binary rather than decimal (as for instance, in sorting out all cards with an X punch in a certain column) an additional factor of four favors the magnetic device.

Taking into account the considerations of the preceding paragraph, it appears reasonable to state that there would be very few if any sorting problems which could be done as fast on the card machine as on the magnetic wire device. For all sorting problems involving a large volume of data, the card machine would probably be slower because of the time lost in manual manipulations, and would certainly be slower when the cards were not filled or nearly filled, or when binary or non-decimal sorting was sufficient.

Although in most cases considerably less than the total capacity of a card is needed to record all of the information which is to be sorted as a unit, problems do arise in which the capacity of a single card is insufficient. If more complicated coding systems are used to enable more information to be put on the same card, new sorting difficulties usually are introduced. On the other hand, if two cards are used instead of one to obtain the needed capacity, the sorting and handling problem is doubled. In contrast to this, no complications or new procedures are introduced in the magnetic wire sorting method by lengthening the information groups. The only result of having larger "units" of information to sort is to lengthen the sorting time in proportion. Thus cases may occur for which the carrying of slightly more information through the sorter will double

CONFIDENTIAL

the card sorting time but hardly affect the wire sorting time.

4.5.4 Systematic Sorting

In the previous sections, the material to be sorted was assumed to be in random order. Sorting machines are often used for processing systematic data. A simple example would be the transposition of a non-symmetrical matrix. A more complicated example occurs in the numerical solution of partial differential equations with three independent variables--- say, two space variables and one time variable. The calculations for the mesh-point $(x, y, t + \Delta t)$ demand information concerning not only the previous results for point (x, y, t) but also the spatial neighbors of this, such as $(x + \Delta x, y + \Delta y, t)$ and $(x - \Delta x, y - \Delta y, t)$, etc. If the internal memory capacity of the computing machine is not large enough to retain all such prior data for many mesh-points, then the external magnetic wire memory must store this information. The various quantities which are needed for calculating the solution at $(x, y, t + \Delta t)$ will originally have been recorded on quite different parts of the wire, but in some systematic way. A systematic rearrangement, or sorting, may be desirable in order to assemble the quantities needed for the next step. However, it will usually be found that with proper use of four input-output devices, there will be no need to rearrange the material on the wires. But if, for any reason, such rearrangement must be done, it can be done more quickly and automatically using wire than using cards.

4.5.5 Collating

The device described in Section 4.5.2 as a binary sorter can function also as a binary collator. One output wire can record information which has been selected from one or the other of two input wires according to any desired plan. A systematic collation might require that N words be

CONFIDENTIAL

taken from one input, the M from the other, then N from the first, and so forth. There is obviously no difficulty in doing this. Often, however, a monotonic collation is needed. In this case, each input tape contains data in monotonic sequence, and it is desired to "mosh" these to form a single monotonic sequence in the output. Each group of words to be transferred as a unit contains a "key word" according to which the various units are ordered. The collator must compare the key words from the two inputs and determine which of the two units of input data is to be recorded next on the output wire. A comparing circuit (see Section 7.6.6) is able, within one or two minor cycles, to determine which of two numbers is the larger. This time is negligible relative to the time required for reading or recording one word on a wire.

Hence any sort of collating, systematic or monotonic, can be carried on at the speed with which the magnetic wire device can transfer information, that is, at a rate of 5,000 or possibly 10,000 pulses per second. Card collating machines are not nearly as rapid as card sorting machines. Moreover, they are binary while sorting machines have as many as thirteen pockets. In comparing sorting speeds, four binary runs on the magnetic wire device were needed to accomplish what is done with one run on the card machines, but in comparing collating speeds, one run on the magnetic wire device accomplishes exactly the same thing as one run on the card collator. It therefore appears that, taking into account only the rate at which information is transferred, the magnetic wire collator must be at least five times faster than the card collator. If the cards are incompletely filled, or if the wire device can be developed to work at more than 5000 pulses per second, the speed ratio might become ten or twenty.

CONFIDENTIAL

The discussion of sorting and collating speeds given in the various subsections of 4.5 has been concerned with magnetic wire devices to handle volumes of data which far exceed the high-speed memory capacity of the EDVAC itself. It should be emphasized again that, when the data to be sorted or collated exist entirely within this high-speed memory, then much higher rates are obtainable.

4.6 SERVO CONTROL FOR MAGNETIC RECORDING

The need for rapidly starting and stopping magnetic wire or tape was brought out in Section 4.5.2. Although devices for doing this have not so far been developed, since commercial sound recording does not require this, servo motors are now obtainable with suitable performance characteristics for this purpose. It is expected that a small motor, with high torque-inertia ratio, will be used to drive that part of the wire which is passing through the recording, reading, and erasing positions. Other servo motors, which need not have as rapid response, can then drive the spools in such a way as to maintain approximately the desirable amount of "slack". It is estimated that the time required to bring the wire from zero speed to recording speed might be one or two milliseconds or even less.

4.7 TYPEWRITERS

Auxiliary devices must be provided for preparing the magnetic records which are to operate the EDVAC, and for automatically printing or typing the output data which the EDVAC records on magnetic wire or tape. Automatic typewriters are now commercially obtainable which will serially type 12 to 18 characters per second and these could be adapted so as to read from magnetic wires or tapes. Higher typing rates are now obtained by multiple equipment, as in the gang printing business machine tabulators which, working from cards, print 80 to 100 characters simultaneously, and

CONFIDENTIAL

do this more than once a second. If equipment is to be multiplied by parallel operation, however, it appears more reasonable to use a number of serial typewriters, each operating from its own serial wire or tape, rather than develop a complicated single gang printing device. This makes the output typewriter more like the input typewriter, too, since the input typewriter is manually operated and must of necessity be a serial device.

It is planned that the input typewriter should have a keyboard somewhat similar to the usual office typewriter, and would type on paper at the same time that it records magnetically. Certain coding operations would be performed by the typewriter. The operator would type numbers in the decimal system, but the typewriter would record these on the wire in a pulse code using four binary digits for each decimal digit. (See Section 6.2). Various letters on the typewriter keyboard would record these in some binary code upon the wire. Possibly an alphabetic code can be provided, to be called into operation by the typewriter shift key.

Bearing in mind the figures given in Section 4.6 relative to starting and stopping the wire, it is evident that each pulse group corresponding to a single decimal digit cannot be placed on the wire directly unless the wire is "backed up" between key operations. (Blank spaces on the wire not only waste wire, but increase the time required later to read off the information.) Although this mode of operation is possible with servo controls, an alternative is suggested. The various small pulse groups corresponding to the individual digit or order keys may be "assembled" in an auxiliary memory, either on a magnetic loop or disk, or in an acoustic tank. When a word is completed, or perhaps a set of words, the entire pulse pattern thus stored can be recorded on the magnetic wire. The wire has to be started and stopped only once for each word, or less often, and the time and wire space required to do this is no longer important.

CONFIDENTIAL

Similarly, the automatic typewriter to produce printed material from data on magnetic wire should have an auxiliary memory, either magnetic or acoustic, so that the wire may be read in units no smaller than a word.

Input typing can be checked by visually proofreading the paper copy which the input typewriter produces, but a more satisfactory method will probably be to have two typists and let each one type the same data. Then the EDVAC can be used to compare automatically the two magnetic wires, and note all discrepancies. Such errors would be recorded on an EDVAC output wire, fed into an output typewriter, and a typed list of the errors to be corrected would then be available.

CONFIDENTIAL

PART 5 BRIEF DESCRIPTION OF EDVAC COMPONENTS

5.0 GENERAL STRUCTURE OF THE EDVAC

The discussion of Part 2 has already indicated many of the characteristics which the EDVAC should have. It should be electronic and digital, and operate serially. It should have a large memory capacity, any part of which can be used for the storage of program information or numbers, including function tables. A central computing device must have immediate or almost immediate access to any part of the memory. Control circuits to direct the computing process must have similar access to the memory. Magnetic wire or tape devices will provide for input and output.

A brief description of the principal components will now be given. The number of such components, and suggested arrangements, are to be found in Part 7, which presents tentative plans for the EDVAC.

5.1 DELAY LINE MEMORY DEVICES (TANKS)

The basic requirement of a large capacity, high-speed memory is met by the use of many acoustic tank units. The operation of a tank has already been described in Part 3. Physically, each acoustic delay line will probably consist of a tube of mercury, one square inch or less in cross-section, with quartz piezoelectric crystals at each end. All such units

CONFIDENTIAL

CONFIDENTIAL

will be mounted within the same thermal chamber for common temperature control. For a nine digit decimal number, at microsecond pulse rates, each short tank will be about two inches long. The long tanks might be as much as six feet long. However, if this length is inconvenient, it is possible to "fold" the tanks by use of reflection so that a long tank will require only three feet.

5.2 ELECTRONIC SWITCHES

If 64 long memory tanks are used, a 64 position switch will be needed to connect a particular one of these tanks to the transmission trunk. A similar switch (or possibly two switches) will be needed to connect the short tanks to the transmission trunk. This switch must have 64 positions and must be able to change from one position to another and open the corresponding gates in the space between two numbers or orders. Such switches can be built to operate in about one microsecond. In the above discussion the space between numbers was given as two microseconds, giving a safety factor of two.

A switch of this type is illustrated in Figure 7. It consists of a resistance matrix with three double entries to the table and eight outputs. The presence or absence of each of three pulses is detected by this arrangement and one of the eight outputs is energized.

The switch for the long tanks would not necessarily have to operate in one microsecond but since two such fast switches must be built perhaps the third should be identical to those both for construction and maintenance reasons.

The controller will probably contain a 32 position switch which chooses the operation to be performed.

CONFIDENTIAL

5.3 COMPUTER

The computer is to perform the operations of addition, subtraction, multiplication and division. In the case of the binary code the computer may perform square rooting with very little added equipment; on the other-hand, in the coded decimal system the process of square rooting is much more complicated. Another operation which must be performed is that of comparison. The process of comparison consists of two parts: first, two numbers are to be subtracted; and second, depending upon the sign of the result the EDVAC is to do one of two things, that is, it is to obey one word of instruction or another. The subtraction part of this process will certainly be done in the computer but where the act of choosing the next order takes place is open to question. The computer will be discussed in detail in Section 7.6.

5.4 CONTROLLER

The controller will receive the orders or words of instruction (from either long or short tanks) and then will cause the EDVAC to carry out that particular instruction. It will consist of a switch which chooses the operation to be performed (depending upon the order) and will signal other switches to operate and what positions to choose (determined by other parts of the order word, see Section 6.2).

5.5 MAGNETIC INPUT AND OUTPUT DEVICES

The input and output devices will read data from magnetic tapes and print results back onto other magnetic tapes. Numbers will be represented by pulses (magnetic pulses) on these tapes. A coded decimal system will be used on the tape regardless of which (coded or binary) is used in the EDVAC proper, because this leads to simplification of the typewriter devices.

CONFIDENTIAL

Not only the pertinent numbers but also all instructions to the EDVAC will be automatically read into the machine from these tapes.

There will be several tape reading mechanisms. Each will have a reading magnet, a clearing magnet, and a recording magnet. That is, each unit can either read from the wire, or record on the wire, and can clear. Possibly, these units will be designed always to clear the wire and if it is desired to keep a record on the wire, the recording magnet will be fed from the reading magnet, so that the wire will receive the same data after passing the clearing magnet. An incidental property of this arrangement is that the deterioration of records because of repeated use is impossible.

The number of tapes should be at least four. This would enable sorting to be done on the tapes themselves, as described in Sec. 4.5. This report will assume that there are to be four tapes.

5.6 OSCILLOSCOPE FOR GRAPHICAL OUTPUT

An oscilloscope and circuits will be provided which will enable the operator to see the result of the computations. This process can consist in part of looking at the numbers stored in some tanks. Of more importance will be circuits which will enable three dimensional vectors to be projected on the screen, the three components being represented respectively by horizontal and vertical displacement and by intensity. In this manner the approximate answer to the problem can be seen whether it consists of a curve or a surface or something else. This picture can change as the computation proceeds. If desirable these circuits can be arranged to show two dimensional sections or projections of a three dimensional figure.

CONFIDENTIAL

There is a question of where to attach this unit, whether to the general transmission trunks or to classify it as a short tank. Another question is: Should there be three channels for the three components of the vector input or should the three components come in serially over the same channel?

5.7 POWER SUPPLIES

Very little can be said about designing the power supplies until decisions have been reached about the rest of the machine. Since it is estimated that there will be about 2,000 tubes in the whole EDVAC the question of the power supply for the EDVAC is trivial compared to that for the ENIAC. Furthermore, since almost all circuits will be a-c coupled, very few voltage levels will be needed and two or three power supplies will suffice.

5.8 AUXILIARY DEVICES

The most important auxiliaries to the EDVAC proper are the means for preparing magnetic wire or tape records by typing from a manuscript, and the means for typing tabulations or other output data automatically from magnetic records. These are described in Section 4.7. A single input or output typewriter will probably not suffice if full use is made of the EDVAC computing capacity. The optimum number will, of course, depend on the type of problems which predominate. Since these devices are separate units, additional ones may be built as required.

CONFIDENTIAL

Various testing devices will also be required for service and maintenance. As in the case of the ENIAC, these can be designed better after the EDVAC has been constructed.

Another auxiliary which may ultimately be found useful, but which is not included in the present plans, is a magnetic wire binary sorter-collator which can function independently of the EDVAC. Such an auxiliary would be able to relieve the EDVAC of large sorting and collating problems, but would not be justified unless the EDVAC were fully occupied with computational problems.

PART 6 NUMBER AND ORDER CODES

6.0 GENERAL CONSIDERATIONS

As indicated in section 2.9.4 the efficiency of the number and order code is a major consideration in designing the machine. That is, the ease with which a problem may be arranged for computation and the desired instructions placed on magnetic wire or tape is a very important factor in the design of the EDVAC.

The choice of operations which are to be represented in the code should be in accordance with the discussion in 2.9.4. Furthermore, the code symbols used should be chosen in accordance with the following principles:

(a) That all the symbols used should be letters or numbers appearing on a standard typewriter. If this is so then anyone anyplace (who knows how) can encode a problem and send the instructions to the place where the machine is located. There, a "typist" can transcribe the coded problem onto the magnetic wire or tape and the problem can be put on the machine with a minimum of time and effort.

(b) That, where possible, the letters denoting some operation be chosen from those which are naturally associated with the operation; for example, "A" for "add".

(c) That the symbols used should not require excessive use of the shift key in the process of encoding a problem.

6.1 LENGTHS OF WORDS AND NUMBERS

The lengths of major and minor cycles (see 3.3) depends upon the order and number codes used, as well as upon the methods used in the computer.

If a binary system is used, a nine digit number can be represented

by 32 pulses (this gives 2^{32} possible numbers). Two more pulses are needed, one to denote the sign and one to distinguish the number from an order (word of instruction). Allowing two pulse times between words this gives 36 pulse times or 36 microseconds as the length of a minor cycle.

In a coded decimal system four pulses are necessary to represent each digit (actually combinations of four pulses can represent 16 different things so there is a lack of efficiency here): Thus, a nine digit number requires 36 pulses. Four more pulses make a minor cycle 40 pulse times or 40 microseconds long.

The length of a major cycle is not subject to as definite a decision as that of a minor cycle. By methods to be explained below, multiplications (using a coded decimal system) can be arranged which take three minor cycles per digit for nine digit multiplication plus 2 minor cycles for preliminaries. Thus, under this system it would seem that a major cycle should be at least this long, and possibly 30 minor cycles. Division, on the other hand, will take (under a similar system to be explained below) six minor cycles per place plus certain preliminary steps taking say, four minor cycles. Thus, for nine place division the time is about 53 minor cycles. Since it is expected that division will not occur frequently in most problems (or at least the operator can arrange for this by using reciprocals, et cetera), there seems to be no great harm in the fact that a division lasts more than one major cycle. The major cycle should not last too long, on the other hand, since time would then be wasted in getting particular numbers or orders out of long tanks into short tanks. Thus for a coded decimal system it seems that the long tanks should be somewhat more than 30 minor cycles long. For

CONFIDENTIAL

CONFIDENTIAL

the purpose of this discussion let us suppose they are 32 minor cycles long. Thus, one long tank will hold 32 numbers or orders.

In the binary system a multiplication or division each take about 33 minor cycles (if performed in completely serial fashion). Thus, in this case the long tanks should be slightly longer, say 34 or 35 minor cycles. Another factor involved here is that lengthening the long tanks increases the memory capacity of the machine with relatively small increase of cost. Thus, there is a conflict between longer tubes for more memory and shorter tubes for better access to numbers and orders.

6.2 CODED DECIMAL SYSTEMS

The following table gives two possibilities for coded decimal systems.

TABLE 2

Digit	Zero Code	Excess three Code
0	0000	0011
1	0001	0100
2	0010	0101
3	0011	0110
4	0100	0111
5	0101	1000
6	0110	1001
7	0111	1010
8	1000	1011
9	1001	1100
unused	1010	1101
	1011	1110
	1100	1111
	1101	0000
	1110	0001
	1111	0010

CONFIDENTIAL

A primary requirement (for sorting purposes) of such a code is monotonicity. Both of the above codes have this property, in fact, if one is added to any number (in a binary manner) the next number is obtained.

An important property that the excess three code has and the other does not is the ease of taking complements. Any complement with respect to nine can be obtained by interchanging ones and zeros in the code, that is pulses are replaced by blanks and blanks by pulses.

The advantage of the excess three code comes with the consideration of addition. If digits are added to digits in binary fashion the answer may turn out to be one of the unused symbols. In this case the code representing the answer has an excess of six instead of three. On the other hand, if there were a carry over beyond the fourth place (that is, the sum of the two numbers were more than 9), the resulting digit would have no excess and three shall be added to it. An adder has been designed which takes care of all these details (see Fig. 6).

CONFIDENTIAL

6.3 THE ORDER CODE

At this point, the general scheme now contemplated for the order code will be outlined. As a practical matter, questions concerning what should be included, and what should not be included, in the system of orders depends upon how many long tanks and how many short tanks are available, what time durations are expected for various operations, and the importance of those various operations in calculating. The decision as to how many pulses make up one word also has a direct effect on the order code, by limiting the information which may be put within one order.

Pulse codes for orders will be taken up in Section 7.3 in conjunction with certain specific suggestions as to switching and control methods. The type of orders which are considered desirable, and which such pulse codes ought to be capable of representing, are given in the list below. This plan for orders is essentially that which von Neumann has proposed after trying out various coding methods on typical problems. His notation, and an equivalent one which uses only ordinary typewriter symbols (in accordance with Section 6.0) are given in parallel columns.

CONFIDENTIAL

Code Symbols (von Neumann) (Suggested)		Meaning
z	z	Denotes a long tank
$\bar{x}$	x	Denotes a short tank
$\rightarrow$	t	Transfer to
$\Rightarrow$	u	To be replaced by
$z \rightarrow \bar{x} / r$	$ztxn$	Transfer r consecutive numbers in tank z into the tanks $x, x+1, x+2, \dots, x+r-1$. If this exhausts tank z then move on to tank $z + 1$
$\bar{x} \rightarrow z / r$	$xtzn$	Transfer r consecutive numbers stored in short tanks $x, x+1, \dots, x+r-1$, into the tank z and if necessary into tank $z+1$.
$c \rightarrow z$	ztp	Connect control to long tank z
$c \rightarrow \bar{x}$	xtp	Connect control to short tanks x .
$x_1 w x_2$	$x_1 w x_2$	Send the numbers in short tanks x_1 and x_2 to the computer and perform the operation w (see below).
$\rightarrow \bar{x} / r$	txn	Transfer the "next" r words into short tank $x, x+1, \dots, x + r-1$.
$\rightarrow z / r$	$tz n$	Transfer the next r quantities into tank z .
$c \rightarrow \bar{x}$	qtx	Dispose of the computer result to short tank x .
. The operations denoted by w are		
$+$	a	Addition
$-$	s	Subtraction
$\cdot$	m	Multiplication
$\div$	d	Division
s	c	Compare (determine which of two numbers is larger, and make some choice of orders to be carried out, depending on the outcome)
$\sqrt{\quad}$	r	Square root (possibly this operation will not be included in the fundamental one above, but will be a subroutine stored in the memory)

CONFIDENTIAL

Von Neumann has specified that some order symbols be capable of modification by deleting a given part of the order and inserting something else in place of this part. This allows for the substitution referred to in Section 2.9.4, so that function tables may be used, subroutines called in, etc.

Several necessary symbols are missing from the above list: these are concerned with ordering the input and output devices to operate.

From examination of the above list in connection with particular pulse code possibilities, it appears that the designation of an operation (exclusive of designating x and z numbers) will require selection from more than 16 but less than 32 possibilities. Thus only 5 pulses would be necessary to make such a selection.

PART 7 TENTATIVE PLANS FOR THE EDVAC

7.0 NUMBER AND ARRANGEMENT OF COMPONENTS

A tentative plan for the EDVAC is illustrated in Fig. 3. As represented on this drawing the machine consists of the units listed in Table 3.

Another possible arrangement is Fig. 4, the interesting feature of which is the fact that all the units are on terminals of the switches which choose the memory tanks. In contrast to these plans the machine could consist of 64 long tanks, 32 short tanks, and the various other units all connected to a trunk system and the necessary isolation provided by a number of gate tubes.

CONFIDENTIAL

Table 3

List of component parts of the EDVAC and tube requirements .

Number of units	Name	Purpose	Estimated number of vacuum tubes
64	Long memory tanks	To store numbers and orders	650
32	Short memory tanks	To store in an accessible manner numbers and orders	350
4	Wire or tape recorders	For input and output of numbers and orders	150
1	Oscilloscope	For visual indication of results of the computations	50
1	Computer	For performing the arith- metical operations	125
1	Control (and switch)	For reading the orders and causing the other units to perform properly	150
1	Power supply (and failure protection)	To supply power	25
3	Electronic switches (64 position; 2 "in and out" and 1 "out") and trunk line amplifiers	To connect the various units to one another	100
Total			1600
If built-in checking, then double the computer and control and add 50			325
Total			1925

CONFIDENTIAL

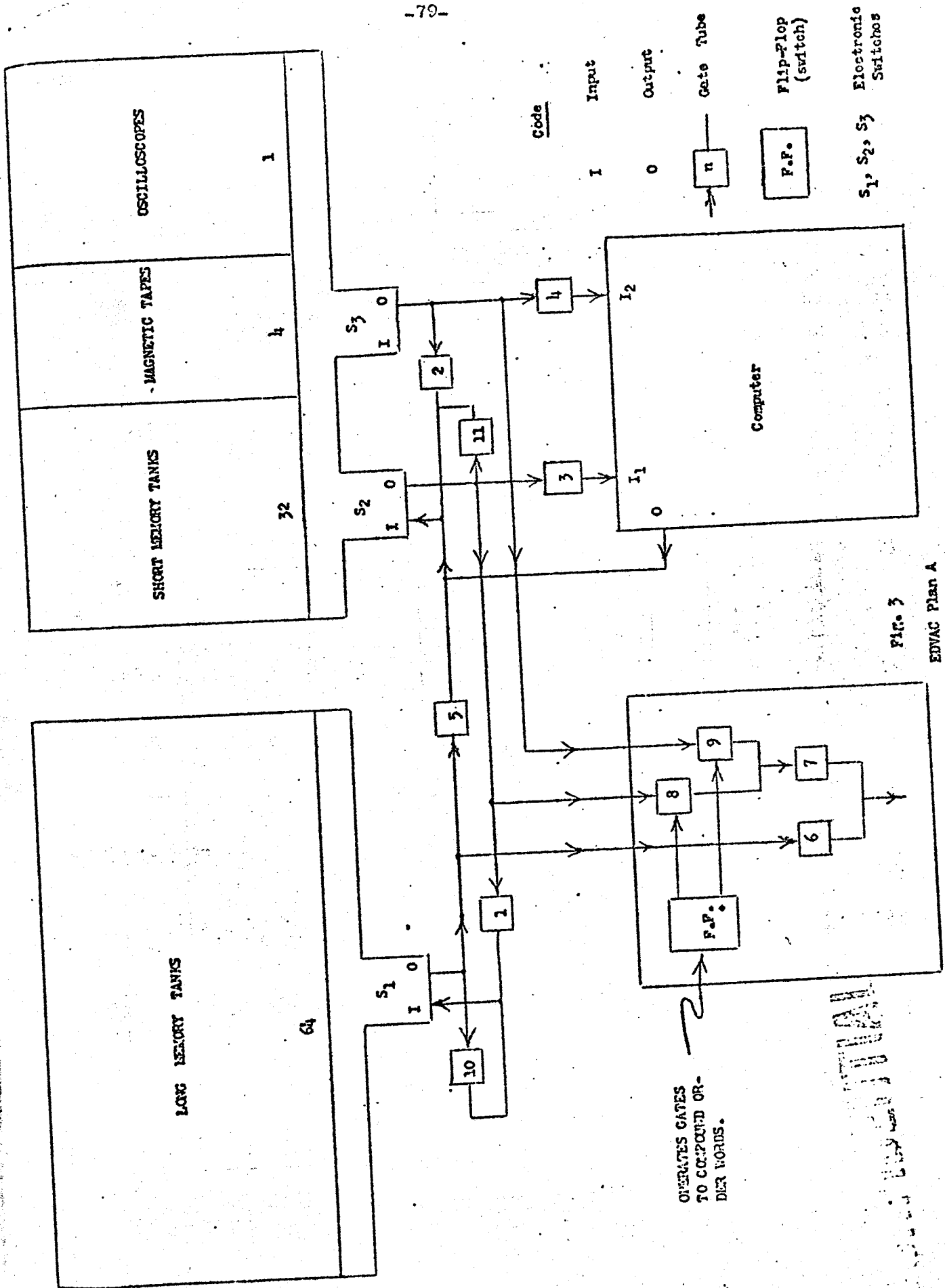
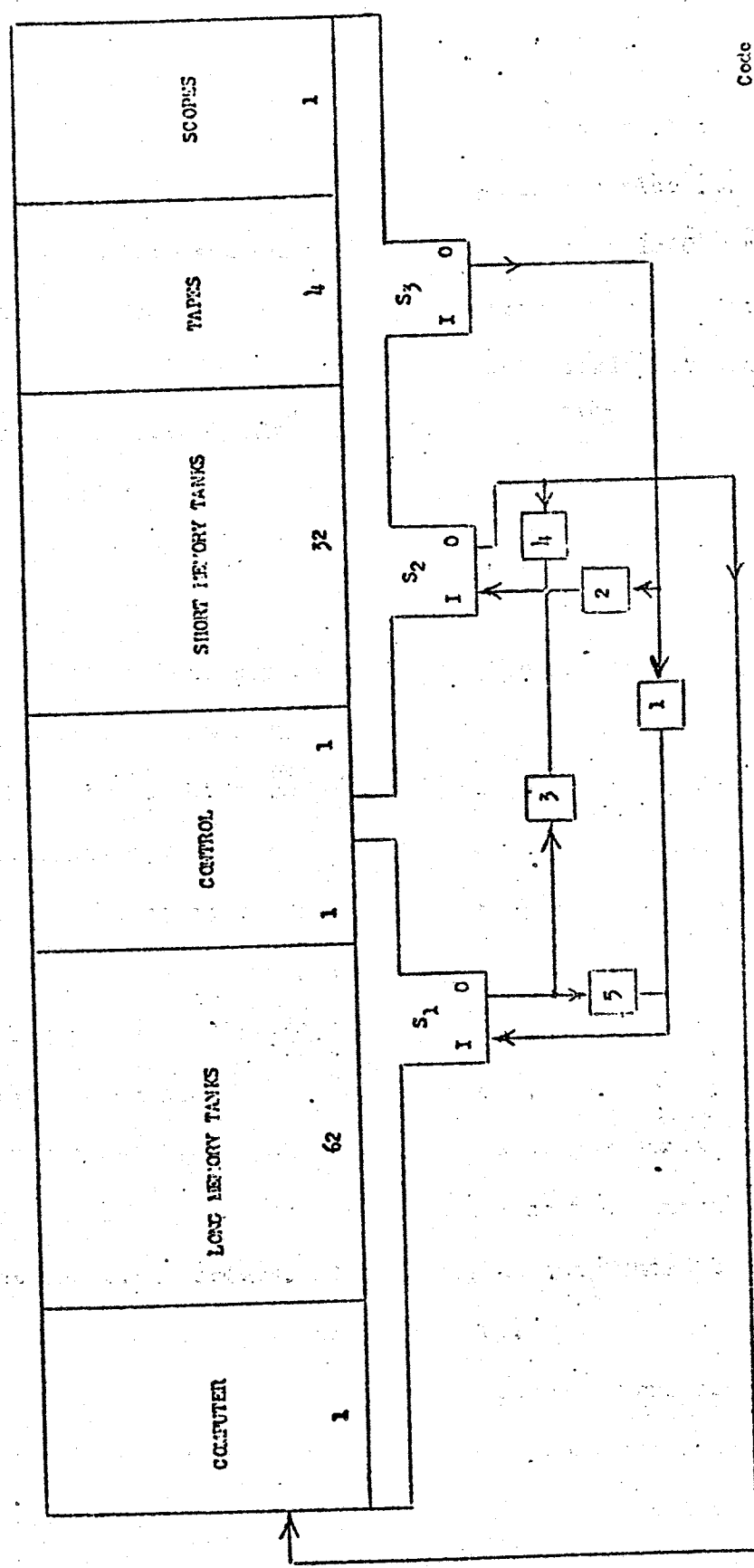


Fig. 3
EDVAC Plan A

OPERATES GATES
TO COMPUTED OR-
DER WORDS.

CONFIDENTIAL



Code	n	Gate Tube
S_1, S_2, S_3	Electronic Switches	
I	Input	
O	Output	

FIG. 4

EDV/C Plan 3

CONFIDENTIAL

The numbers of the various kinds of units listed above have been arrived at by considering a few of the problems which the machine is expected to solve. If the long tanks are 30 minor cycles (or words) long the operator has on tap here once each major cycle over 1800 words. Combining this with the almost unlimited memory capacity of the magnetic tape (even though the numbers are not available here so quickly) it seems that very few problems will exceed this capacity.

The short tanks are provided so as to have numbers available each minor cycle. Also, they are convenient in sorting problems. In an example worked out by von Neumann (a problem which consists of "meshing" in a monotone manner two monotone sequences) some 25 short tanks are needed to store all the pertinent instructions of the process, while about 32 more are needed to be able to shift numbers from one long tank into another which may already contain some numbers. That is, the numbers are first placed into small tanks then at the proper time they are transferred into some other long tank. The transfer directly from one long tank to another is not desirable because of the difference in phase which would necessitate the leaving of spaces (as many as 29 spaces).

By using about two more minor cycles per sorting step the instructions can be temporarily stored in a long tank and all the short tanks used in the shifting process. Thus, only as many short tanks as there are minor cycles in a major cycle are necessary.

The number of magnetic tapes is perhaps more subject to revision than some of the other numbers given above. Four seems to be the minimum number in order to be able to do sorting on the tapes themselves. In fact, it might be found very convenient to have at least a fifth tape.

CONFIDENTIAL

7.1 ESTIMATE OF TUBE REQUIREMENTS

Table 3 gives an estimate of the number of tubes required to build the EDVAC. Table 1 gives a comparison between the number of tubes used in the EDVAC and the number used in the ENIAC. The smaller number of tubes required for the EDVAC means that: (1) it is easier to construct, (2) it is easier to maintain, (3) the problem of ventilation is much simpler, (4) the cost of power for operation is less, and (5) the machine takes up less floor space.

Assuming the use of standard size radio tubes it is probable that the EDVAC will occupy a space of less than 180 cubic feet. This could be arranged to be a space two feet by eleven feet by about eight feet high. Allowing room for access to the wiring and the memory tubes as well as to the magnetic tape mechanisms the floor space necessary for the EDVAC would be, say, eight feet by fifteen feet.

If small size vacuum tubes were used, the above estimates of the space required could be reduced considerably.

The 325 tubes listed (in Table 3) for checking purposes would be used as follows: 275 of these would be used to duplicate the control and computer units; the extra 50 tubes would be used to compare the results of the two computers and to compare the two interpretations of each order given by the two control units. If the results or the interpretations did not agree the machine would stop and give a signal to attract the attention of the operator. An oscilloscope could then be used as an aid in finding the trouble.

CONFIDENTIAL

7.2 SUGGESTED SWITCH CONNECTIONS

Table 4 gives a possible arrangement of the connections on the switches for the set up illustrated in Fig. 4, (Plan B).

Table 4

Suggested Switch Connections

No.	Terminal Code	S ₁ (In and Out)	S ₂ (In and Out)	S ₃ (Out only)
0	000000	Control	Control	----
1	000001	Tank 1	Tank 1	Tank 1
...				
32	100000	Tank 32	Tank 32	Tank 32
33	100001	Tank 33	Tape 1	Tape 1
34	100010	Tank 34	Tape 2	Tape 2
35	100011	Tank 35	Tape 3	Tape 3
36	100100	Tank 36	Tape 4	Tape 4
37	100101	Tank 37	Oscilloscope	----
38	100110	Tank 38		
..	...	...		
63	111111	Computer		

In the case of the arrangement illustrated in Fig. 3 (Plan A), the difference in connections is as follows: Terminal 0 (code: 000000) on all switches goes to tanks 0, and terminal 63 (111111) on S₁ goes to long tank number 63.

CONFIDENTIAL

7.3 PULSE CODE FOR ORDERS

Just what circuits are in the controlling unit depends upon what arrangement is used. With that illustrated in Fig. 4 the controlling unit can be relatively simple as will be shown below. In the case of Fig. 3 it seems best to have a 32 position electronic switch in the controlling unit which chooses the particular operation to be performed. Whether all 32

positions will be used is somewhat doubtful, but more than 16 will certainly be needed. There will be a difference in the type of order code used in the two cases. The following table gives a code list and the pulse representations in the two systems.

TABLE 5 Order Codes

von Neumann's Code	Alternate Code	Plan A		Plan B	
		Pulse Code	Switches and Gates*	Pulse Code	Switches and Gates*
$z \rightarrow \bar{x} / r$	ztxn	00001	$S_1 S_2$ [5]	---	$S_1 S_2$ [3]
$\bar{x} \rightarrow z / r$	xtzn	00010	$S_1 S_2$ [1]	---	$S_1 S_3$ [1]
$C \rightarrow \bar{x}$	xtp	00011	S_3 [7] [9]	---	$S_1 S_3$ [1]
$C \rightarrow z$	ztp	00100	S_1 [6]		$S_1 S_2$ [3]
$O \rightarrow \bar{x}$	qtx				
$\rightarrow \bar{x} / r$	txn				
$\rightarrow z / r$	tzn				
$x_1 + x_2$	$x_1 a x_2$	00101	$S_2 S_3$ [3] [4]	001	$S_1 S_2 S_3$ [1]
$x_1 - x_2$	$x_1 s x_2$	00110	"	010	"
$x_1 \cdot x_2$	$x_1 m x_2$	00111	"	011	"
x_1 / x_2	$x_1 d x_2$	01000	"	100	"
$x_1 S x_2$	$x_1 c x_2$	01001	"	101	"
(Reading)	ftzn	01010	$S_1 S_2$ [1]		$S_1 S_2$ [1]
(Recording)	ztfn	01011	$S_1 S_2$ [5]		$S_1 S_3$ [3]
(Viewing Scope)	ztvn	01100	$S_1 S_2$ [5]		$S_1 S_3$ [3]

* Gates not listed in this table must be closed. Possible code words for the two cases considered above will now be given.

Plan A

1 00001 00110 000011 000111 0000001 Total pulses = 30
 (1) (2) (3) (4) (5) (6)

(1) The first pulse denotes that this word is an order and not a number. Numbers would start with a blank.

000000000000000000

(2) Referring to the above table, this is the code for transferring from a long tank to a short tank.

(3) This symbol determines what word is picked out of the tank chosen by switch S_1 .

(4) This is the instruction for the setting of switch S_1 . The numbers here range from 0 to 63.

(5) This denotes the setting of the switch S_2 .

(6) This number is r . In either a binary or decimal coded system the words would be long enough to make this number larger than the 128 allowed in the above order.

The above order would cause word number 6 in long tank 3 to be transferred to short tank 7.

Plan B

1	000001	000011	000110	001001	001	000	Total pulses = 31
(1)	(2)	(3)	(4)	(5)	(6)	(7)	

Here the various groups of pulses cause the following operations:

(1) Order distinction.

(2) Setting of switch S_1 .

(3) Word chosen from long tank, or with one more pulse in the case of arithmetical operations this can determine the disposal.

(4) Determines the setting of switch S_2 .

(5) Determines the setting of switch S_3 .

(6) Determines the arithmetical operation to be performed by the computer.

(7) These pulses take care of certain other details, such as whether the tapes run forward or backward, etc.

CONFIDENTIAL

CONFIDENTIAL

In this case the number r has not been indicated. Values of r up to 128 would require seven more pulses and would make the code word about 33 pulses long which is consistent with the length of the numbers in the coded decimal system. Thus, this code is a little long for use in the binary system unless more than nine digit numbers are to be used.

From here on the discussion will center around Plan A since the code is more efficient and there is possibly more freedom in the use of the units. On the other hand the design of case two is somewhat simpler since the control unit only has to route the groups of pulses to the proper place instead of choosing from 32 possibilities.

7.4 CONTROL

With the code of Plan A the controlling unit must contain a 32 position switch which chooses the transfer or arithmetical operation to be performed. The order designation allows the order to get into the controlling unit. The first five digit code causes the control switch to set up. The group which indicates the word to be chosen from a long tank must be used to cause the rest of the order to circulate in a short tank (probably located in the control itself) the proper number of times before the switches S_1 , etc., are set up.

One possible solution for determining the proper number of times of such circulation is to have a binary adder located in the control unit which adds -1 to the word number once each circulation. The switching process can be arranged to take place when this number reaches zero or perhaps at the time of the first overdraft (that is, when the number becomes -1 itself).

This binary adder will serve another purpose. In case of trans-

CONFIDENTIAL

CONFIDENTIAL

fering r numbers the last parts of the order must circulate until the proper word (or space, depending upon the direction of transfer) is available at the long tank. Then the switch S_1 must connect to that tank and remain connected for r minor cycles (or possibly change to the next long tank if the first one is exhausted) and the short tank switch must change from tank to tank (that is, it must change over a sequence $x, x + 1, x + 2, \dots, x + r - 1$). This changing can be accomplished by the adder adding 1 to the number of the short tank one each circulation and adding -1 to the number r . Then the process can be terminated when r becomes zero (or again the code can be adjusted so there is an overdraft to determine the end of the process.

Frequently the machine must combine two different words, usually a number and a word of instruction with blanks in it, to form another word of instruction. This can be done by having a gate arrangement operated by a flip-flop (which can change in between pulses) which choose parts of the two words which are coming in on two channels. (see gates 8 and 9, Fig. 3). The necessity of combining order words occurs in function table processes. Suppose the values of some function are stored in a long tank, or say, several long tanks. Let the argument correspond to the numbers of the tanks and the word positions in the respective tanks. Usually the argument will be the result of a series of computations and thus will be stored in some short tank, say. In another short tank there must be stored an order to transfer from long to short tanks. Whether there are blanks or not in this word for the number of the long tank and the word position seems immaterial. By the type of device described above the control can be connected to the two short tanks and the order to transfer fed into the control. At the proper instant the connection can be changed from one channel to the other

CONFIDENTIAL

and the long tank number and word position read out into control. If three order words are to be combined the process can be carried out in two steps. The result of the first combination could be circulated through the short tank in control and then fed back into the combining unit along with the third word to participate in the combination.

It is clear from the above description that there is need for various timing devices to operate in connection with the circuits of the controlling unit. These can be fed from a central oscillator or clock which causes all the units to operate in synchronism and controlled by various gating arrangements. Some problems of this type can be solved by a judicious use of delay lines. Since the design of the controlling unit has not as yet reached a stable state the discussion of these details will be saved until some later time.

7.5 DELAY LINE MEMORY CIRCUITS

Figure 1 of part 3 gave a simplified block diagram of a delay line memory. Fig. 5 is a more detailed version of the circuits (showing input and output gates) associated with a long or a short tank. Pulses from the central timing clock are fed in through the gate C. From here they go to a tube F which will drive the delay line. Probably the circuit of the tube F will be that of a blocked oscillator. This circuit has the advantage that a relatively weak signal will drive it and since the input is through a transformer a positive or negative signal can be used to turn the tube on.

The output of the tank feeds through a three stage amplifier (A) and a pulse former (which produces a broad pulse to overlap the next clock pulse arriving at C) to gates G, O₁, and O₂. Gate G is normally open and

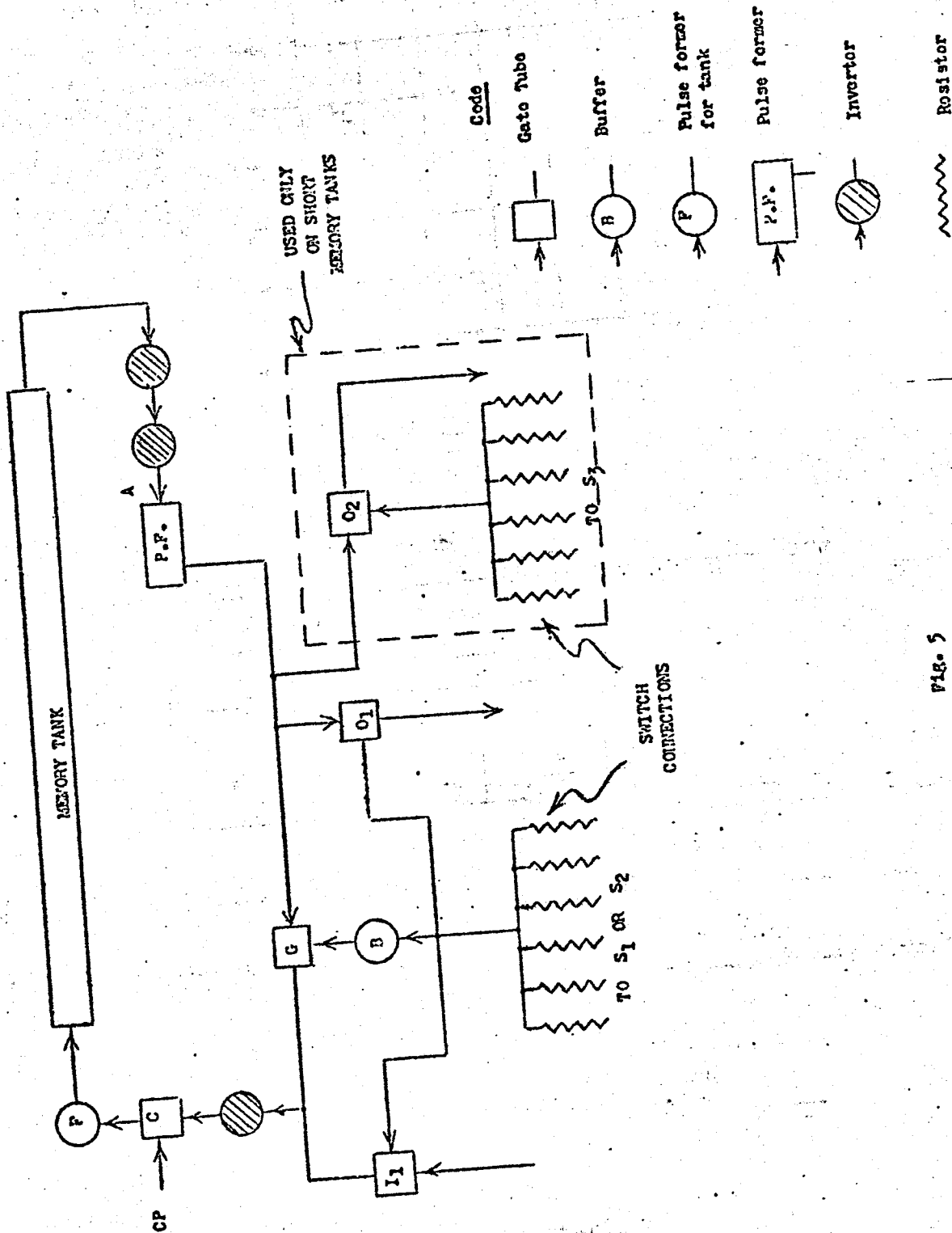


Fig. 5
Memory Tank Circuit

CONFIDENTIAL

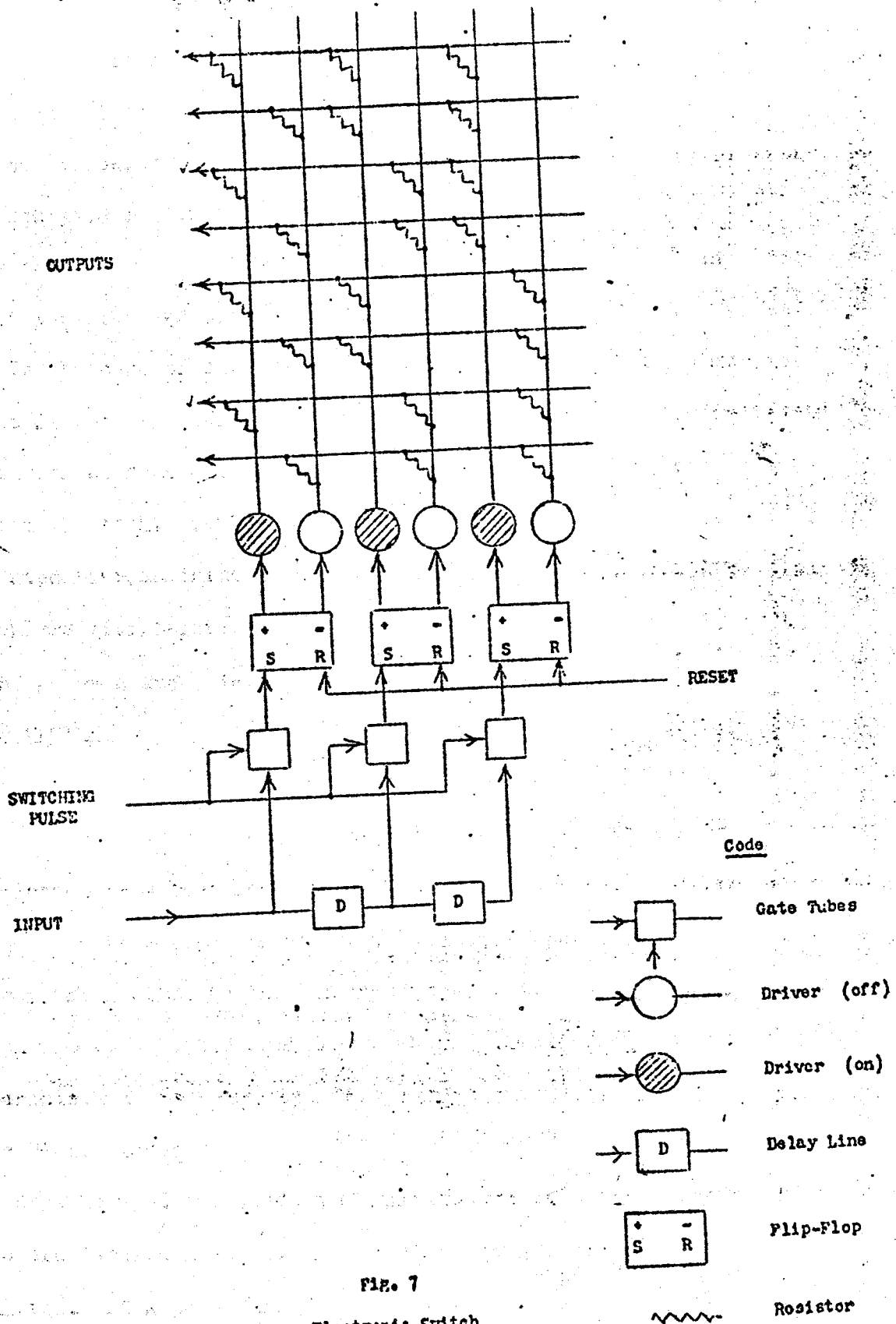


FIG. 7

Electronic Switch

CONFIDENTIAL

allows the pulses to circulate around the memory tank.

When switch S_1 or S_2 is connected to this tank the gates I and O_1 are opened allowing data stored in the tank to come out and new data to go in through gate I. In this case the buffer B causes the gate G to close preventing the circulation of the data and hence insuring that any incoming data is not superimposed upon any that is already in the tank.

In the case of short tanks the output of the pulse former goes out to gate O_2 for a second output. Under the arrangement illustrated in Fig. 5 data can be read out through switch S_3 and the information not be cleared from the tank. On the other hand, if S_2 (or S_1 in the case of long tanks) is used to read information out then the information will be cleared as well, unless recirculated through an external gate. (Gates 10 or 11 in Fig. 3, and gates 4 and 5 in Fig. 4)

7.6 THE COMPUTER

7.6.0 The Computer

The computer required for the coded decimal system is considerably more complicated than that required for the binary system. On the other hand if one adds to the equipment of the binary computer that required to convert from the decimal to the binary system on the input and output (also probably on the line to the oscilloscope) the amount of equipment is probably comparable.

7.6.1 The Binary Adder

Drawing PY-2-101 gives a wiring diagram of a binary adder. The two inputs are labeled A and B. If no pulse arrives on either line nothing is transmitted. If a pulse arrives on one of the lines it is transmitted through tubes 14, 17, and 27. If a pulse arrives on both inputs simultaneously it is transmitted through tube 5 blocking gates 3 and 4. This carry-

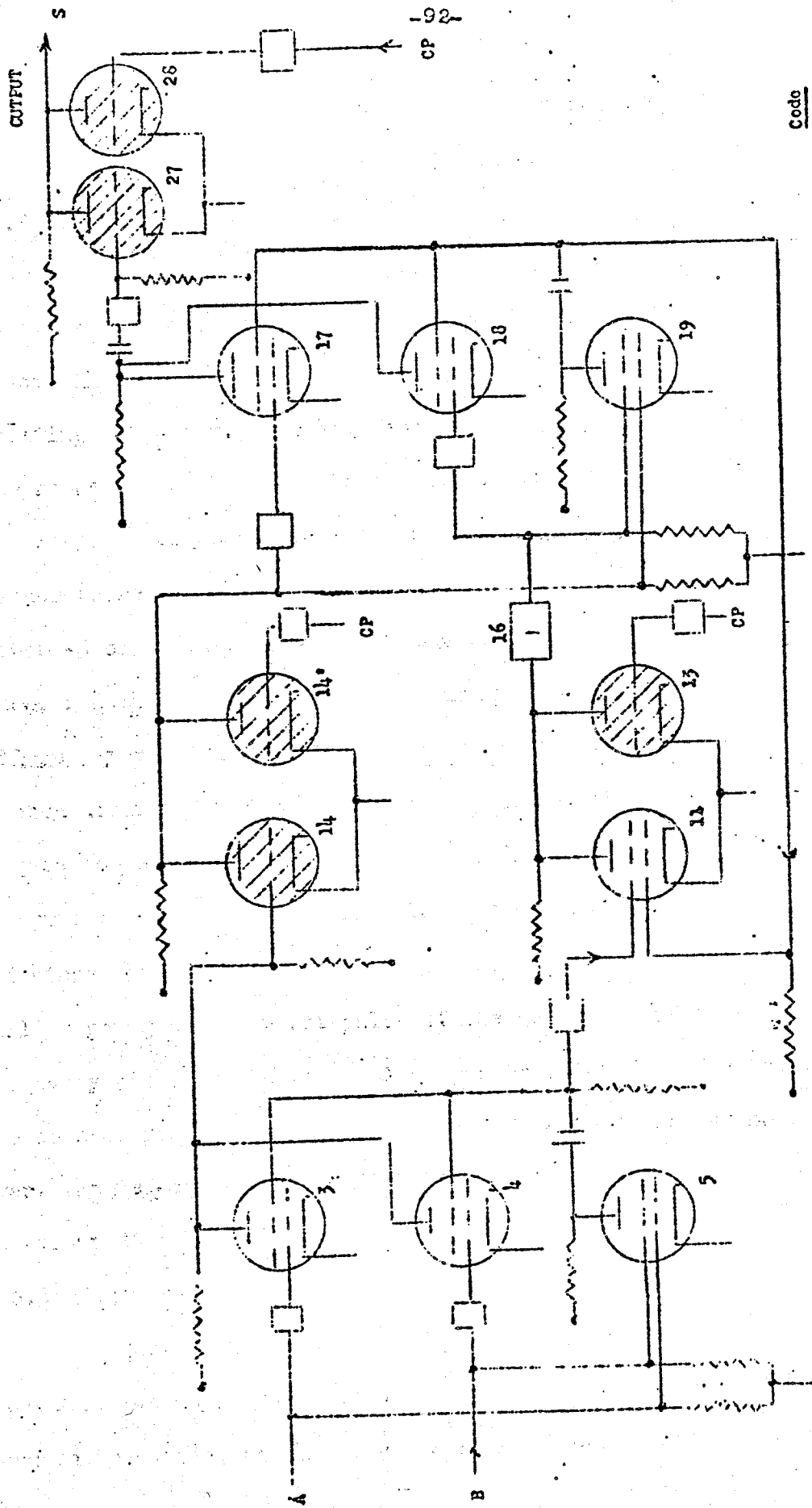


Fig. PY-2-101 (Simplified): Binary Pulse Adder

- Corrective delays
- One pulse-time delay for carry
- Normally conducting tubes

Code

CONFIDENTIAL

over pulse will be properly delayed and added into the next digit by tubes 17, 18, and 19. If this leads to a second carry over, the signal comes from gate 19 and goes back through tube 11.

7.6.2 The Coded Decimal Adder

The following coded decimal adder has been designed to work with the excess three code described in Table 3. A simplified block diagram appears in Fig. 6. This adder consists essentially of two binary adders operating in series. The addition is actually performed in the zero (binary) code rather than the excess three code. If a_3 and b_3 represent digits in the excess three code then the sum will be a number c_0 or c_6 depending upon whether there was a carry or not. Thus to get the proper answer, c_3 , either +3 or -3 must be added to the result. This is the reason for the second binary adder. There is a four pulse counter which very probably will be a three pulse delay line which is fed by a pulse at the beginning of each digit. The first two outputs furnish the pulses 0011 through the +3 gate to be combined with the result from the first adder. The last two give the combination 1101 which represents -3. To operate these plus and minus three circuits the fourth carry over comes out to a flip-flop as well as going to the first pulse of the next digit. This flip-flop is reset every digit by the four pulse counter. The fourth output pulse of the counter goes to the second binary adder to suppress the fourth carry over. It is estimated that the coded decimal adder described here will use about 35 tubes.

7.6.3 Negative Numbers

There are a number of ways of handling negative numbers. It is intended that negative numbers are to be printed in true (not complementary) form on the tapes, to simplify the typewriters.

CONFIDENTIAL

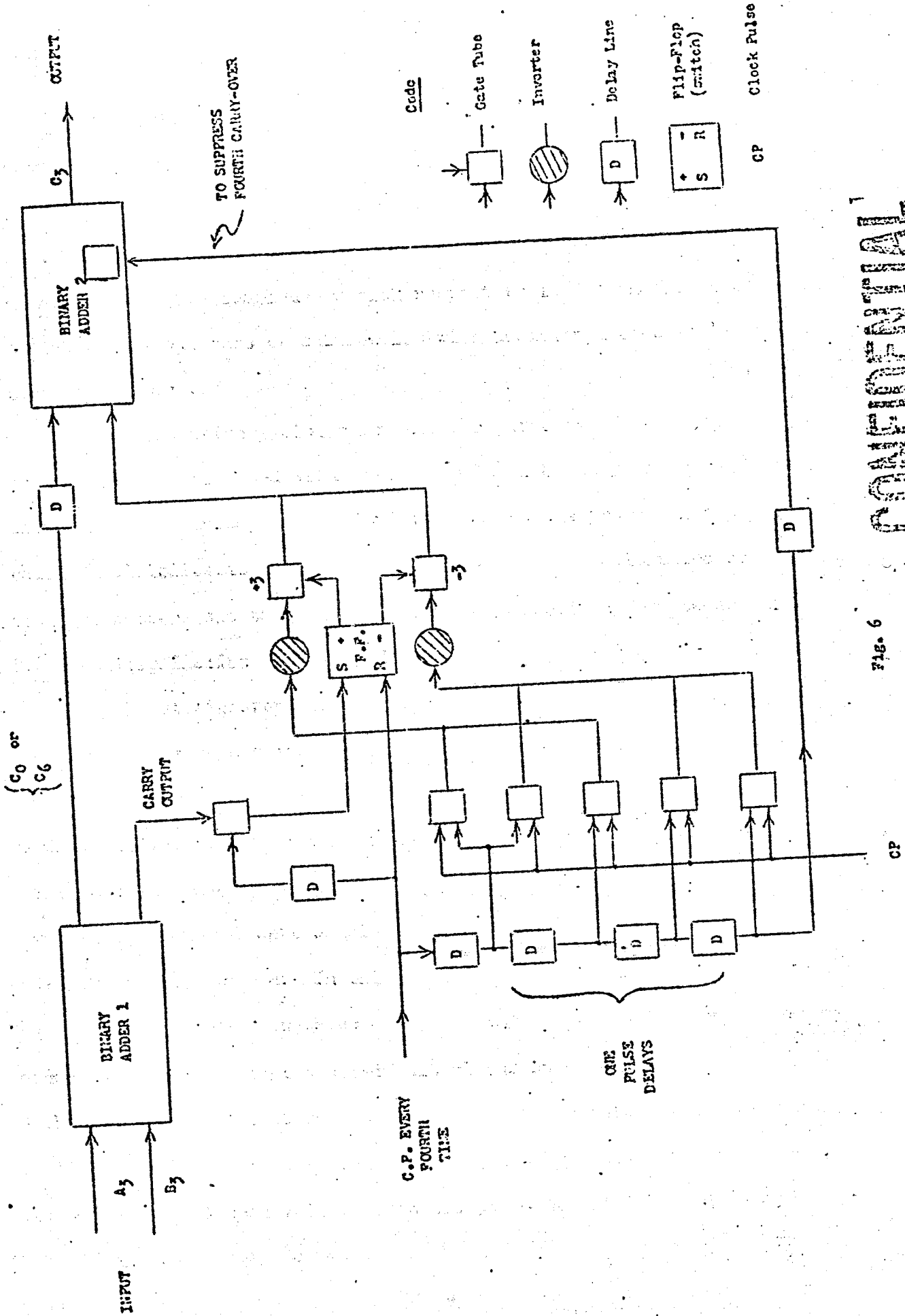


Fig. 6
Decimal Adder

CONFIDENTIAL

Negative numbers can be used in true form throughout the machine provided the computer is furnished with a subtracting circuit. In contrast to this, negative numbers can be complemented (with respect to either 10^n or 10^n-1 ; here serial operation makes it difficult to provide the correction pulse to give complements with respect to 10^n) then the regular adder circuits can take care of subtraction with additional circuits to take care of the sign.

A tentative decision is that for purposes of addition all negative numbers are complemented with respect to 10^n-1 as they enter the adder. The answer is then "uncomplemented" if need be and the proper sign provided. In case of multiplication and division the numbers are multiplied or divided as positive numbers and then the proper sign is provided for the answer.

7.6.4 Multiplication

To attain speed in multiplication two short memory tanks in the computer are used to store the multiplicand and four times the multiplicand. This storage can be accomplished in two addition times which (since the numbers are positive) can be two minor cycles. Thereafter the digits of the multiplier are fed into the computer where they activate an electronic switch which decades what combinations of the multiplicand and four times the multiplicand to use. In this manner, multiplication by any digit can be done in not more than three addition times. If subtractions take two minor cycles each the high speed multiplication described here would require a third short tank to store the complement of the multiplicand with respect to 10^n (this is required since multiplication by the digit 7 requires 4 times the multiplicand twice and subtracting the multiplicand once).

CONFIDENTIAL

7.6.5 Division

The efficient way of doing division consists of storing in the two auxiliary tanks one and three times the divisor. Then division at any particular place consists of subtracting three times the divisor until overdraft and then adding on one times the divisor until another overdraft occurs. In this manner, division at any one place cannot take more than six operations of addition or subtraction. Since the numerator can be treated as a positive numerator the additions can be done in a minor cycle each. Furthermore, if three times the divisor can be prepared as the complement of the true number with respect to 10^n the subtractions reduce to additions of positive numbers (with the last carry over ignored except to indicate overdraft) and can each be done in a minor cycle.

7.6.6 Comparing

In many computing problems, the instructions will at some time depend upon certain numerical results already obtained. The ability to modify instructions, or choose a particular subsequent routine from two or more possible routines, is necessary if the EDVAC is not to be seriously limited in flexibility and generality.

In giving the instruction to compare two numbers, the operator will specify, implicitly or explicitly, where the control unit is to find the next word of instruction, in accordance with the outcome of this comparison. The two numbers will be subtracted, and the sign of the difference will be used to choose between two alternatives. A reasonable way of specifying the alternatives is to have the controller continue in normal fashion if the result is positive, but to specify explicitly where the next order word is to be found if the result is negative. That is, for a

CONFIDENTIAL

positive difference, the next order is to be in the word immediately following the word which ordered the comparison.

The comparing operation therefore makes use of the computer to subtract two numbers, and then, if the difference is negative, switches the input of the controller to whatever position is designated by the comparison order.

7.6.7 The Decimal Point

A great burden is placed upon the person who codes an intricate computing problem if he has to estimate the magnitudes of the numbers which will arise in the course of computation. If all numbers in the machine must fall between -1 and 1 , and can have no more than 9 decimal digits, say, then an operator who can estimate the sizes of the various numbers which will be encountered in his work to within a factor of 100 can (except in subtractions) be sure of about 6 significant figures; if he tries to get more, he runs the risk of producing numbers having the magnitudes greater than unity, which the machine will not handle properly.

This difficulty can be relieved in either of two ways--one can simply provide more significant figure capacity, which the operator can throw away by less careful planning, or one can add an "exponential" system to keep track of powers of 10 (or 2), and cause all numbers to be registered to the same number of significant digits. This latter solution is more efficient from the point of view of efficient use of the memory capacity, but has the disadvantage of complicating the computer, especially in addition and subtraction. This system is sometimes referred to as the "floating decimal point".

The floating decimal point is considered so advantageous from the

CONFIDENTIAL

operational point of view that the additional equipment required for it would be justified. However, for a first machine, the development and design work must be considered as well as the amount of equipment, and in view of this it appears wiser, in making tentative plans, to develop first a "fixed decimal point" machine.

PART 8 RELIABILITY AND CHECKING

8.0 CHARACTER OF FAILURES

Before any reasonable discussion of checking can be undertaken, it is necessary to consider the reliability of the devices being checked. No amount or complication of checking apparatus is of much use if the device being checked is likely to fail often, so that a large percentage of the time is spent in experimentally discovering faults by extremely thorough checking procedures. On the other hand, there is less and less need for checking as the reliability of components is increased. Component failures can never be prevented entirely, however, and in digital computing systems almost any component failure may vitiate the results. Therefore, although each component may have an exceedingly low rate of failure, it is unlikely that a computing machine containing many such components will fail so seldom as to make checking devices unnecessary.

Checking devices detect rather than prevent failures. But they, too, are subject to failure, so that it might seem that there is no end to checking. The aim of a checking device is not, however, to detect every failure, but to reduce the probability of having an undetected failure to some extremely small value which is deemed tolerable. Unfortunately, it is rather difficult to compare or design checking systems because good statistical data on failure characteristics are available for only a few of the components and systems to be compared. The difficulties are enhanced by the fact that engineering prac-

CONFIDENTIAL

tice is poorly standardized in some fields and techniques, and reliability may vary widely in spite of the similarity of the basic components and methods employed.

In an electronic computer, it is worthwhile attempting, first of all, to improve the reliability of the vacuum tubes which are used as components. However, at present it is not considered feasible to have special tubes solely to gain reliability. As a matter of fact, commercial vacuum tubes which are in high production are more likely to be reliable than special tubes produced in small quantities. Rigorous testing and inspection is a better way to improve reliability. Operation of the tubes with currents and voltages which are well under the manufacturer's ratings also will improve lifetime and consequently reliability. There is, however, no large body of experimental data available on the reliability of tubes used under such conditions. This is one of the difficulties already referred to. Finally, a great deal may be done in designing the circuits which use these tubes so that reliability is improved. Methods of construction, and methods of testing the complete assembly, must be such as to facilitate quick repairs when a fault has occurred.

Checking methods may be devised which do more than indicate the existence of a fault--they may, in fact, be designed to help localize the failure and thus be an aid in servicing the machine. To what extent this can be done cannot be considered at this time, but possibilities of this sort will certainly govern final decisions as to what checking methods are employed.

It is convenient to classify failures into two broad classes:

- (1) Lasting failures, which render the apparatus inoperative until the fault is corrected, and

CONFIDENTIAL

- (2) Transient failures, for which the apparatus passes back and forth, at intervals, between an operative state and an in-operative state. These are sometimes referred to as "intermittents".

Type 2 faults are of chief concern, because of the difficulty of detecting and locating such faults. Also, if such faults are not detected immediately, there is no way of knowing how many calculations have been spoiled. This type of fault becomes more insidious as the "failure interval" is decreased, because no appreciable time is allowed to make a test, even assuming that a method exists for telling when the fault occurs.

In complex computing machines, relays have been preferred to vacuum tubes very largely because relays have a relatively high reliability and considerable is known about their failure characteristics. Type 1 faults are extremely improbable for relays. Consequently, about 85% of relay faults are of type 2, which are even more undesirable. In order to meet this problem, checking circuits have been included in some relay machines to insure that no operation can proceed unless the immediately prior operation has been checked. This check is based on a system in which a fixed number of relays must always operate, no matter what operation is being done. Such a checking circuit not only notices the failures, but localizes them. Such techniques have been most widely applied in automatic telephone exchange equipment.

In electronic computing machines reliability is increased by using the tubes as switches (or on-off devices). Most of the failures can be traced to defective tubes, and these failures are usually type 1. Then it is necessary only to detect that such a failure has occurred, and the location of the fault can be discovered by more ordinary means. A checking system analogous to that just described for relays is more difficult

CONFIDENTIAL

to provide with vacuum tubes because of their limited number of terminals.

The question of reliability has not received as much attention in electronic equipment as it has in relay equipment for several reasons. Vacuum tubes have had a shorter development period, and for many of their uses reliability has not been a significant factor. For instance, low cost is considered more important for home radios. Heretofore, no machines using vacuum tubes in large numbers has existed to focus attention on reliability.

Certain general characteristics of vacuum tube failures are worth noting. If a number of tubes are put into service, and all these tubes have already passed certain short period tests (which may reject 5 or 10% of the original lot), then most of the failures for some thousands of hours will be type 1, and most of these type 1 failures will occur during the first few hours of use. The failure rate after these first hours becomes low and remains low for a long time. Ultimately, as the tubes age, their cathode emission falls off to the point where they would no longer pass the original test. This gradual decrease in output is likely to cause type 2 failures, since a weak, borderline signal may sometimes be effective and at other times not. A checking system which has a high probability of detecting such situations may be used, or the likelihood of this type of failure can be reduced by periodically removing all tubes and subjecting them to an "external" test. An intermediate possibility also exists: it may be possible to design the circuits so that, by altering certain supply voltages, tubes which are approaching this borderline operation will fail to operate. Thus, potential type 2 failures are forced into becoming actual type 1 failures. This is, in effect, a way of testing the tubes without removing them from the machine.

Since type 2 failures in relay machines are usually the result

CONFIDENTIAL

of dirt momentarily fouling the contacts, there is no comparable procedure for relay machines.

Power supply failure is of little importance to relay machines, since there is usually but one supply and the failure will probably be evident either because it is type 1 or because many relays will be affected by a type 2 failure. A vacuum tube machine like the EDVAC will have several power supplies and if an extremely short-period transient should affect any one, errors may be produced. Such a transient might be of external origin and occur very infrequently; hence the occurrence of every such transient must be noted. It is not difficult to monitor the various supplies, however, since there are only a few of them. By comparatively simple equipment, all aberrations in the power supplies which might cause trouble can be detected. The checking circuit which does this is independent of all other checking devices, and faults arising from the power supplies are thus localized.

8.1 TYPES AND LEVELS OF CHECKING

It is convenient, in discussing checking methods and devices, to use the following terminology:

- Repeat checking - the same operation is repeated with the same equipment
- Multiple checking - the same operation is done with two or more separate pieces of equipment of identical construction
- Inverse checking - one operation is checked by performing an inverse operation, the result of which is compared with the original data on which the first operation was performed
- Identity checking - two equivalent methods of doing the same operation are employed, and the results compared. The two operations may use entirely different equipment, or the same equipment in different ways.

CONFIDENTIAL

- Smoothness checking - smoothness of a series of values is criterion (usually tested by examination of some order of higher differences)
- Programmed checking - checking which is carried out in accordance with specific instruction orders (at the control of the operator)
- Built-in checking - checking which is not at the discretion of the operator, and which is always done by the machine without instruction
- Continuous checking - checking which will detect any single fault no matter when it occurs.
- Intermittent checking - Periodic or occasional checking, which can detect faulty operation only if such faults occur at specific times (sample checking)

It is also convenient to speak of various "levels" of checking. At the lowest level, every pulse in the memory might be checked, while at the highest level, an entire problem might be multiple checked by having two EDVAC machines and comparing the problem solutions which they turn out. Both of these methods are unduly expensive, since it is expected that other checking methods will insure adequate reliability with less equipment than these require.

8.2 CONTINUOUS VERSUS INTERMITTENT CHECKING

Continuous checking is employed in the relay machines already mentioned. An equivalent method in the EDVAC memory would require that a method be introduced which would distinguish between a zero digit (no pulse) and a failure (which gives no pulse). This could be done by using pulses of opposite sign to represent zero digits, or by using two pulse positions for each binary digit, and requiring that at least one pulse appear but never two for a given digit. Either method would require considerably more electronic equipment, and the second would cut the speed of the machine and necessitate longer tanks (unless memory capacity were

CONFIDENTIAL

to be sacrificed). Continuous checking is necessary for relay machines, however, because of the peculiarity of the type 2 failures to which they are subject. A particle of dust may prevent closure of a contact during one operation, and on the very next operation the contact may close properly. Behavior of this sort is not typical of vacuum tubes; those type 2 failures which might occur by reason of borderline operation can be guarded against by the methods suggested in Section 8.0. In general, intermittent checking is a reasonable method for vacuum tubes, but a poor method for relays.

8.3 CHECKING IN SERIAL MACHINES

Serial operation has been advocated for the EDVAC because by this method less equipment is necessary. This has, of course, an obvious correlation with reliability. Since the failure of any component may be counted as a failure for the whole machine, the probability of machine failure is reduced when the number of components is reduced. The time lost by such failure is also reduced because both diagnosis and repair are easier with less equipment.

Quite apart from this, however, serial operation is advantageous from the point of view of checking. Not only are there fewer components to be checked, but sometimes a number of components can be checked by a single checking device. Checking at a higher level by programming operations to test all components is also simplified, since with fewer components and more of them used in each operation, a programmed check is much easier to work out. Since many components in a serial machine are in almost continuous operation, there is more opportunity for detection and location of type 2 faults. The fact that all of the digits for any number must pass through the same channels is extremely important. This means that a type 1

failure in such a channel will affect all digits of every number which uses that channel, so that even without any checking devices the calculations would be so completely upset as to make the failure obvious when the results were inspected. In contrast to this, failure of a single digit channel in a multi-channel machine may cause only a small error in the results. Such an error might be systematic enough so as to go undetected by differencing (smoothness tests), and would only be revealed by identity or multiple checking or by a programmed test problem. When type 2 failures are considered, the fact that all digits must pass through the same channel is even more important. In a serial machine, any type 2 failure will ultimately cause a gross error, and will be detected without aid of checking devices. In a multi-channel machine a type 2 failure may occur in a digit line without ever causing a gross error, and if the faulty component is used only infrequently it may remain undetected through many problems.

Although all failures in a serial machine are likely to cause gross errors, and therefore would not long escape detection if checking devices were not employed, this is not an argument for omitting checking devices in a serial machine. Any checking device which does not enormously increase the equipment should be built in, and experience with the completed machine can then determine to what extent programmed checking must be resorted to.

8.4 CHECKING IN OTHER MACHINES

It may be of interest to point out that, with the exception of the relay machines which use continuous checking, all digital machines heretofore employed for complex computing problems have relied entirely on programmed checking. On card machines, multiple checking is employed where possible, with some checking done by identity methods. A larger automatically programmed digital machine uses inverse and identity checking which must be programmed.

For analogy machines, test problems, which are a variety of programmed intermittent checking, are customary. No matter what machine is used, results are tested for smoothness whenever the nature of the problem permits. When errors are frequent, locating them by examining difference becomes difficult if not impossible. Furthermore, certain types of systematic errors will not be

detected by a smoothness test.

8.5 CHECKING THE COMPUTER AND CONTROL UNIT

In Section 8.2 it was stated that intermittent rather than continuous checking would be reasonable for vacuum tube memory devices. A continuous check would essentially double the equipment requirements, and this did not seem justified. The amount of equipment involved in the computer and control unit, however, is small enough that duplication of these units is not unreasonable. This would enable continuous checking by the multiple method.

There seems no attractive alternate method. Repeat checking is of slight value --- when the same operation is repeated with the same equipment, type 1 failures will never be detected, and type 2 failures of period longer than two operations may easily escape detection. Identity checking is more useful at a higher level, by combining the elementary operations in different ways and in different order to achieve the same result. The elementary operations of addition and multiplication can be checked to some extent with a single computer by interchange of the two numbers entering the operation. Some component failures may nevertheless remain undetected by this method, so that it is not a complete check. And no interchange of this sort is permissible in division nor possible in square root. For these operations, inverse checking could be used --- that is, division could be checked by multiplication, and square root by squaring, followed by a comparison with the original dividend or radicand. The additional control equipment to make this check built-in might be considerable. Also, all of these alternative methods at least double the computing time, since the checking cannot be simultaneous with the direct operation when only one computer is available.

CONFIDENTIAL

The only alternative to which the above objections do not apply is to omit built-in checking, and rely on programmed checking. To this, however, a most serious objection exists, since it would increase the burden upon the operator who must plan and code the problem. Further, programmed checking at the level of the elementary operations is very wasteful of time; the computing time might easily be trebled. Programmed checking at a higher level would be less wasteful of time, but would fail to localize faults.

To check computer operations within any standard operation such as multiplication or division would be possible, but seems unnecessary. Computer and control failures will be sufficiently localized if checking is done by comparing the results from duplicate computers after completion of a standard operation.

It is therefore proposed that two identical controls and computers be provided for the EDVAC. Further, each should be provided with a comparison circuit. At the end of a standard operation, each computer transmits its result to the comparison circuit of the other, so that the comparisons are also carried out in duplicate. Both comparison circuits must show computer agreement for the check to be complete. Thus a tube failure in a comparison circuit cannot go undetected; if it could, computer errors might not be discovered when they first occurred, and there would be no way of knowing how many calculations were spoiled.

Having duplicate computer, control and comparison units is also an advantage for service and maintenance purposes. It is unlikely that both would fail at the same time, or in the same way. Hence, diagnosis of a failure in one can be aided by direct observation of the normal performance of the other. Test data which could not reasonably be included in even a large volume of service history is thus available for maintenance purposes.

8.6 CHECKING THE MEMORY AND SWITCHES

There is fundamentally only one good way of checking a memory and that is to have duplicate memories, with a comparison circuit (or duplicate comparison circuits as in the computer) for indicating disagreement. This is essentially multiple checking, and doubles the equipment. The probability of duplicate errors in such a system will be extremely small, since the only common equipment is the power supply, which is already monitored. The continuous checking methods suggested in Section 8.2 are somewhat similar, but are more like repeat checking. Essentially, the two memory systems are then in the same tanks and share some of the equipment. Certain failures might cause a tank to be filled either with "zeros" or "ones" independently of the actual correct input, and the checking system would not detect this.

A method which has a high probability for noticing errors and which does not strongly exhibit the fault just mentioned is one which provides, in coded form, a number in each tank which gives the sum of the digits of the other numbers in that tank, or for the coded decimal system, this number might be the total number of pulses which appear in the rest of the tank. This number must be properly altered each time a transfer is made. The difficulty with this system is that the checking system itself becomes rather complicated.

One might also make use of the fact that, in the excess-three binary decimal code (Section 6.2) every digit symbol contains at least one pulse and at least one blank. The order code would then have to have a similar property. This provides a certain check against type 1 failures and has a fair chance of catching some type 2 failures. It is probably no better than other systems now to be discussed which have the additional advantage of checking the switches.

A compromise method which provides an intermittent check on the memory and a continuous check on the switches is to store "tags" in the various tanks. The tag stored in each tank is the number which corresponds to the switch position associated with that tank. A single comparison device, associated with the switch and controller, serves as the checking circuit. Whenever the switch selects a particular tank, the output of that tank is connected to a trunk line (through some amplifier circuits). At the beginning of each major cycle, the tag for the tank to which the switch is set is thus on this trunk line, and may be put into the comparison circuit. Simultaneously, the number which set the switch is fed to this comparing circuit and disagreement indicates a failure. This failure might be in the switch, in the tank, or in the amplifier circuits between the tank and the trunk line. All of these are checked by this method. Only type 2 failures of duration less than a major cycle would escape detection by this system. In order to check the input trunk line, amplifiers and input tank gate, it is sufficient to require that the tank contents be regenerated through this circuit, rather than through the recirculation gate at the tank itself. Detection of switch failures which result in connecting two tanks simultaneously may be accomplished by arranging that the superposition of two tags is always different from either of the tags which are superimposed.

While the above method is reasonable for long tanks, the inclusion of a tag in short tanks would increase their length, and this in turn would require the long tanks to be increased ^{proportionately} in length, which is undesirable. An intermittent check of the short tanks can be arranged by using the time during which the switch operates to check the presence of a standardized tag (alike for all short tanks) consisting of one pulse and one space. In this case, the switch would be checked independently by circuits which would verify that it was set up in accordance with the number sent it.

8.7 CHECKING THE INPUT AND OUTPUT

The best check on the input and output is, as in other components, obtained by multiple checking. With four magnetic devices, two may be used for input and two for output, so that this method may be employed by programming duplicate operation. In view of the small amount of electronic equipment which these units contain, this extreme measure does not seem warranted. Moreover, since both positive and negative magnetization will probably be used (for reasons explained in Section 4.3) these may be used for a built-in check on the transfer to and from wire. Faulty operation of the servo devices will be evident without elaborate checking devices.

Duplicate checking does seem to be the preferred method of eliminating errors which the typist may make in preparing magnetic wire records. Two different typists, using different machines if possible, should type the same material. Their duplicate records may then be compared by the EDVAC or some auxiliary device, and all discrepancies transferred to another wire. This wire is then used to operate the automatic typewriter which produces a printed list of the discrepancies.

8.8 SMOOTHNESS CHECKING AND PROGRAMMED CHECKING

With the various built-in checking methods which the EDVAC will have, it is expected that the undetected failures will be very rare. If this expectation is not fulfilled, or if the peculiarities of some problem require even greater guarantees against error, additional checking may be programmed. This option has in no way been limited by the checking devices which are built in.

Smoothness checking, for instance, may be necessary even when no failures occur, because of limited accuracy and rounding-off errors. In such cases, the EDVAC can compute the differences required for this check. If

desired, these may be scanned by means of the visual output oscilloscope.

PART 9 CONCLUSION

At the time this report appears, most of the general principles underlying the design of the EDVAC have been considerably clarified by discussions which were carried on concurrently with work on the ENIAC. However, decisions on many specific questions raised in the course of those discussions must await the results of experiments which are now being undertaken. The designs presented in this report are therefore merely illustrative examples, and should not be considered final.

It is planned to issue supplementary reports to cover topics which were necessarily omitted from this one. In particular, it is intended that some of these shall present examples of problems coded in accordance with proposed instruction codes.

CONFIDENTIAL

